Interactive Journal of Medical Research

Impact Factor (2022): 2.0
Volume 14 (2025) ISSN 1929-073X

Contents

Corrigenda and Addenda

Correction: Digital Health for Australia: Bridging the Rural, Regional, and Remote Health Gap (e89755)
Shakeel Mahmood, M Huda, Kedir Ahmed, Thapa Subash, Feleke Astawesegn, Anayochukwu Anyasodor, Mohammad Moni, Muhammad Shiddiky, Utpal Mondal, Setognal Aychiluhm, Santosh Giri, Allen Ross. . . . 9

Viewpoints

The Impact of Online Labor Platforms on Workforce Management in Health Care (e68546)
Maryam Ahmadi Shad, Michael Simon, Florian Liberatore. . . . 11

Systemic Inflammation and Disruption of the Local Microenvironment Compromise Muscle Regeneration: Critical Pathogenesis of Autoimmune-Associated Sarcopenia (e64456)
Yingjuan Zhang, Qingqian Wu, Yi Wang, Qingyan Chen, Shuang Han, Bei Li, Qingwen Zhao, Qianzhuo Wang, Yule Wang, Yue Gao. . . . 19

How to Design Electronic Case Report Form (eCRF) Questions to Maximize Semantic Interoperability in Clinical Research (e51598)
Eugenia Rinaldi, Caroline Stellmach, Sylvia Thun. . . . 33

The Evolution of Uroflowmetry and Bladder Diary and the Emerging Trend of Using Home Devices From Hospital to Home (e66694)
Ming-wei Li, Yao-Chou Tsai, Stephen Yang, Yuan-Hung Pong, Yu-Ting Tsai, Vincent Tsai. . . . 40

The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations (e59823)
Dimitri Roustan, François Bastardot. . . . 48

An Automated Clinical Laboratory Decision Support System for Test Utilization, Medical Necessity Verification, and Payment Processing (e46007)
Safedin Beqaj, Rojeet Shrestha, Tim Hamill. . . . 55

Use of Clinical Public Databases in Hidradenitis Suppurativa Research (e70282)
Xu Liu, Linghong Guo, Xian Jiang. . . . 61

XSL•FO
RenderX

Effective Recruitment or Bot Attack? The Challenge of Internet-Based Research Surveys and Recommendations to Reduce Risk and Improve Robustness (e60548)
Liesje Donkin, Nathan Henry, Amy Kercher, Mangor Pedersen, Holly Wilson, Amy Chan. 65

Digital Health for Australia: Bridging the Rural, Regional, and Remote Health Gap (e67460)
Shakeel Mahmood, M Huda, Kedir Yimam Ahmed, Thapa Subash, Feleke Astawesegn, Anayochukwu Anyasodor, Mohammad Moni, Muhammad J A Shiddiky, Utpal K Mondal, Setognal Aychiluhm, Santosh Giri, Allen Ross. 77

Reviews

Acupuncture and Moxibustion for Poststroke Depression: Systematic Review (e76577)
Lu Meng, Chuang-Long Xu, Xiao-Xu He, Xiao-Chan Tan. 87

Adoption of Telemedicine for Dementia Care in Nigeria: Scoping Review (e75168)
Abiodun Adedeji, Huseyin Dogan, Festus Adedoyin, Michelle Heward. 106

Obesity Prevention and Reduction in China Using the Social Media Platform WeChat: Scoping Review (e65538)
Yinuo Wang, Xuxiu Zhuang, Samantha Sundermeir, Joel Gittelsohn. 119

Developing Requirements for a Standardized System to Return Individual Research Results Back to Study Participants: Narrative Review (e65606)
Rosalyn Carr, Vita Chan, Nicholas West, Matthias Görges. 139

Economic Evaluations and Equity in the Use of Artificial Intelligence in Imaging Examinations for Medical Diagnosis in People With Dermatological, Neurological, and Pulmonary Diseases: Systematic Review (e56240)
Giulia Santana, Rodrigo Couto, Rafael Loureiro, Brunna Furriel, Luis de Paula, Edna Rother, Joselisa de Paiva, Lucas Correia. 154

Digital Interventions for Improving Body Dissatisfaction in Children and Emerging Adults: Systematic Review and Meta-Analysis (e72231)
Li Liu, Jianning Yang, Fengmei Tan, Xia Yang, Huan Luo, Yanhua Chen, Xiaolei Zhao. 170

Horizon Scanning Methods for Health Care Technology Innovation Identification: Rapid Scoping Review of Patent Research Studies (e70323)
Sonia Garcia Gonzalez-Moral, Erin Pennock, Olushola Ewedairo, Elizabeth Green, James Elgey, Andrew Mkwashi. 188

eHealth Literacy Interventions: Scoping Review (e69640)
Yan Wang, Yutian Niu, Rongjing Xu, Qingqing Zhang, Shoumei Jia, Anni Wang. 206

Safety and Efficacy of Aspirin and Indobufen in the Treatment of Atherosclerotic Diseases: Systematic Review and Meta-Analysis (e75363)
Wenhao Pan, Linger Guan, Haicheng Zhang. 220

Reasons for Crown Failures in Primary Teeth: Systematic Review and Meta-Analysis (e57958)
Stephan Lampl, Deepa Gurunathan, Deepak Mehta, Krithikadatta Jogikalmat. 235

The Effect of Combining mHealth and Health Professional–Led Intervention for Improving Health-Related Outcomes in Chronic Diseases: Systematic Review and Meta-Analysis (e55835)
Masashi Kanai, Takahiro Miki, Toshiya Sakoda, Yuta Hagiwara. 255

XSL•FO
RenderX

Comparing Digital Versus Face-to-Face Delivery of Systemic Psychotherapy Interventions: Systematic Review and Meta-Analysis of Randomized Controlled Trials (e46441)
Pieter Erasmus, Moritz Borrmann, Jule Becker, Lars Kuchinke, Gunther Meinlschmidt. . . . 268

Global Evidence on the Sustainability of Telemedicine in Outpatient and Primary Care During the First 2 Years of the COVID-19 Pandemic: Scoping Review Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) Framework (e45367)
Daniela Valdes, Ankit Shanker, Ghofran Hijazi, Daniel Mensah, Tahir Bockarie, Ioana Lazar, Siti Ibrahim, Hamid Zolfagharinia, Rob Procter, Rachel Spencer, Jeremy Dale, Armina Paule, Liam Medlin, Keerthana Tharuvara Kallottil. . . . 289

Therapeutic Guidelines for the Self-Management of Major Depressive Disorder: Scoping Review (e63959)
Priscila de Campos Tibúrcio, Priscila Maria Marcheti, Daniela Miori Pascon, Marco Montebello Junior, Maria Alzete de Lima, Carla Fernandes, Célia Samarina Vilaça de Brito Santos, Maria Socorro de Sousa Nóbrega. . . . 311

Comparison of Outcomes Between Staged and Same-Day Circumferential Spinal Fusion for Adult Spinal Deformity: Systematic Review and Meta-Analysis (e67290)
Mert Dagli, Ryan Turlip, Felix Oettl, Mohamed Emara, Jaskeerat Gujral, Daksh Chauhan, Hasan Ahmad, Gabrielle Santangelo, Connor Wathen, Yohannes Ghenbot, John Arena, Joshua Golubovsky, Ben Gu, John Shin, Jang Yoon, Ali Ozturk, William Welch. . . . 331

Informatics Interventions for Maternal Morbidity: Scoping Review (e64826)
Jill Inderstrodt, Julia Stumpff, Rebecca Smollen, Shreya Sridhar, Sarah El-Azab, Opeyemi Ojo, Brendan Bowns, David Haggstrom. . . . 345

Innovative, Technology-Driven, Digital Tools for Managing Pediatric Urinary Incontinence: Scoping Review (e66336)
Lola Bladt, Jiri Vermeulen, Alexandra Vermandel, Gunter De Win, Lukas Van Campenhout. . . . 365

Ethical Challenges and Opportunities of AI in End-of-Life Palliative Care: Integrative Review (e73517)
Abel García Abejas, David Geraldes Santos, Fabio Leite Costa, Aida Cordero Botejara, Helder Mota-Filipe, Àngels Salvador Vergés. . . . 388

Guideline-Based Digital Exercise Interventions for Reducing Body Weight and Fat and Promoting Physical Activity in Adults With Overweight and Obesity: Systematic Review and Meta-Analysis (e73656)
Mohamad Motevalli, Clemens Drenowatz, Derrick Tanous, Gerhard Ruedl, Werner Kirschner, Markus Schauer, Thomas Rosemann, Katharina Wirnitzer. . . . 402

Cross-Platform Availability of Smartphone Sensors for Depression Indication Systems: Mixed-Methods Umbrella Review (e69686)
Johannes Leimhofer, Milica Petrovic, Andreas Dominik, Dominik Heider, Ulrich Hegerl. . . . 426

Integration of Conventional and Virtual Reality Approaches in Augmented Reality for Theory-Based Psychoeducational Intervention Design for Chronic Low Back Pain: Scoping Review (e59611)
Robin Conen, Steffen Mueller, Ana Tibubos. . . . 515

Process Evaluations of Interventions for the Prevention of Type 2 Diabetes in Women With Gestational Diabetes Mellitus: Systematic Review (e51718)
Iklil Sa'id, Natasha Hotung, Madeleine Benton, Iliatha Nadal, Anisah Baharom, Matthew Prina, Barakatun Nisak Mohd Yusof, Kimberley Goldsmith, Samantha Birts, Ching Siew Mooi, Angus Forbes, Khalida Ismail, Boon Chew. . . . 844

Original Papers

Glucagon-Like Peptide-1 Receptor Agonists Combined With Personalized Digital Health Care for the Treatment of Metabolic Syndrome in Adults With Obesity: Retrospective Observational Study (e63079)
Hala Zakaria, Hadoun Jabri, Sheikha Alshehhi, Milena Caccelli, Joelle Debs, Yousef Said, Joudy Kattan, Noah Almarzooqi, Ali Hashemi, Ihsan Almarzooqi. . . . 442

XSL•FO
RenderX

Interactive Effects of Weight Recording Frequency and the Volume of Chat Communication With Health Care Professionals on Weight Loss in mHealth Interventions for Noncommunicable Diseases: Retrospective Observational Study (e65863)
Yuta Hagiwara, Takuji Adachi, Masashi Kanai, Kotoe Shimizu, Shinpei Ishida, Takahiro Miki. . . . 454

Application of Machine Learning and Emerging Health Technologies in the Uptake of HIV Testing: Bibliometric Analysis of Studies Published From 2000 to 2024 (e64829)
Musa Jaiteh, Edith Phalane, Yegnanew Shiferaw, Lateef Amusa, Hossana Twinomurinzi, Refilwe Phaswana-Mafuya. . . . 465

The Effects of Virtual Reality on Hope and Travel Expectations in Healthy and Hospitalized Children: Quasi-Experimental Design Approach (e65311)
Pei-Shan Hsieh, Hsiu-Sen Chiang, Fang-Liang Huang, Wen-Hsin Chi. . . . 484

Mapping Hot Spots and Global Research Trends in Exergaming Between 1997 and 2024: Bibliometric Analysis (e66738)
Abdullah Kayhan, Elif Kose, Burcu Kayhan, Nan Zeng. . . . 498

Using an Electronic Immunization Registry (Aplikasi Sehat IndonesiaKu) in Indonesia: Cross-Sectional Study (e53849)
Dewi Aisyah, Astri Utami, Fauziah Rahman, Nathasya Adriani, Fiqi Fitransyah, M Endryantoro, Prima Hutapea, Gertrudis Tandy, Logan Manikam, Zisis Kozlakidis. . . . 527

Size and Composition of Caregiver Networks Who Manage Medications for Persons Living With Dementia: Cross-Sectional Analysis of the 2011-2022 National Health and Aging Trends Study (e64499)
Reed Bratches, Frank Puga, Paul Barr, Amanda Leggett, Meredith Masel, James Odom, Rita Jablonski. . . . 545

Determining the Prioritization of Behavior Change Techniques for Long-Term Stroke Rehabilitation: Delphi Survey Study (e59172)
Agata Wróbel, Philip Cash, Anja Maier, John Paulin Hansen. . . . 555

Community Comfort With Automatic Sharing of Race, Ethnicity, and Language Data Between Health Care Settings: Cross-Sectional Study (e67288)
Noah Brazer, Baylah Tessier-Sherman, Deron Galusha, Sakinah Suttiratana, Corrine Liu, Katherine Kim, Mark Abraham, Marcella Nunez-Smith, Karen Wang. . . . 568

Evaluating the Safety and Efficacy of a Non–Weight-Centric Approach to Obesity Prevention in Rural and Urban Female Adolescents: Quasi-Experimental Study (e71341)
Areeg Zuair, Rola Jalloun, Naif Alzahrani, Fahad Alhowaymel, Esraa Merza, Bandar Alhumaidi, Mohammad Alahmadi. . . . 580

Simplified Medication Adherence Questionnaire (SMAQ) for People Living With HIV in a National Hospital in Mexico: Instrument Validation Study (e59562)
Luis Del Moral Trinidad, Luz González Hernández, Jaime Andrade Villanueva, Pedro Martínez-Ayala, Adriana Valle Rodríguez, Vida Ruíz Herrera, José Vizcaíno Résendiz, Melva Herrera Godina, Sergio Dominguez-Lara. . . . 601

Interstep Variations of Stairways and Associations of High-Contrast Striping and Fall-Related Events: Observational Study (e60622)
Sara Harper, Chayston Brown, Shandon Poulsen, Tyson Barrett, Christopher Dakin. . . . 610

Understanding Loneliness Through Analysis of Twitter and Reddit Data: Comparative Study (e49464)
Hurmat Shah, Mowafa Househ. . . . 618

XSL•FO
RenderX

Exploring 97 Years of Aedes aegypti as the Vector for Dengue, Yellow Fever, Zika, and Chikungunya (Diptera: Culicidae): Scientometric Analysis (e65844)
Yoon Cheong, Sumarni Mohd Ghazali, Mohd Mat Hashim, Mohd Che Ibrahim, Afzufira Amran, Tsye Tiunh, Hui Lim, Yong Cheah, Balvinder Gill, Kuang Lim. . . . 632

Characterization of Global Research Trends and Prospects on Prone Positioning in Respiratory Failure: Bibliometric Analysis (e67276)
Rong Lei, Feng Yue, Chaofu Yue, Zihan Zhang, Xian Huang, Qiaolin Li, Zhigang Yang, Rong Li, Keyi Zhao, Mei Yang. . . . 653

Lactation Consultant Access and Breastfeeding Outcomes in the United States: Cross-Sectional Analysis (e70098)
James Dockins, Heather Pahl, David Lingerfelt. . . . 670

Impact of the COVID-19 Pandemic on Contraceptive Services at Selected Primary Health Care Facilities in India, Nigeria, and Tanzania: Cross-Sectional Study (e59874)
Rita Kabra, Beena Joshi, Ester Elisaria, Tanimola Akande, Komal Allagh, Adesola Olumide, Deepti Tandon, Ranjan Prusty, Mary Ramesh, Donat Shamba, Bhavya MK, Shabana Khan, James Kiarie. . . . 684

Student and Physician Views of How the Dobbs Decision Affects Training and Practice Location Preferences: Cross-Sectional Questionnaire Study (e55035)
Morgan Levy, Simone Bernstein, Sarah McNeilly, Abigail Liberty, Shira Fishbach, Shikha Jain, Jessica Gold, Vineet Arora. . . . 697

Readmission After Ischemic Stroke in Ningxia, China, From 2017 to 2021: Retrospective Cohort Study (e67522)
Hua Meng, Xingtian Wang, Dongfeng Pan, Xinya Su, Wenwen Lu, Zhuo Liu, Yuhui Geng, Xiaojuan Ma, Ting Pan, Peifeng Liang. . . . 709

Exploring Barriers to Patients' Progression in the Cardiac Rehabilitation Journey From Health Care Providers' Perspectives: Qualitative Study (e66164)
Shri Ramesh, Darwin Jull, Hélène Fournier, Fateme Rajabiyazdi. . . . 722

Mapping the Dynamics of Inhibitors and Facilitators of Exercise Behavior Within the Transtheoretical Model: Nationwide Cross-Sectional Study Using Text Mining Analysis (e77400)
Kosuke Sakai, Kota Fukai, Yuko Furuya, Shoko Nakazawa, Kei Sano, Masayuki Tatemichi. . . . 735

Long-Term Engagement of Diverse Study Cohorts in Decentralized Research: Longitudinal Analysis of "All of Us" Research Program Data (e56803)
Vijay Yadav, Elias Neto, Megan Doerr, Abhishek Pratap, Larsson Omberg. . . . 750

Dengue Epidemiology in 7 Southeast Asian Countries: 24-Year, Retrospective, Multicountry Ecological Study (e70491)
Shun-Long Weng, Fang-Yu Hung, Sung-Tse Li, Bo-Huang Liou, Chun-Yan Yeung, Yu-Lin Tai, Yi-Hsuan Wu, Ya-Ning Huang, Nan-Chang Chiu, Liang-Yen Lin, Hsin Chi, Chien-Yu Lin. . . . 763

Factors Predicting Information Overload During the COVID-19 Pandemic in the Digital Age: Longitudinal Study (e67098)
Hiroko Okada, Tsuyoshi Okuhara, Takahiro Kiuchi. . . . 771

Factors Influencing Physicians' Referral Decision-Making for Rehabilitation Outpatient Services in the Health Care Landscape of China: Cross-Sectional Study (e64464)
Yawei Li, Ruixue Ye, Linlin Shan, Kun Wang, Kaiwen Xue, Zeyu Zhang, Yingzi Hao, Yucong Zou, Xiaoxuan Li, Yulong Wang. . . . 784

Policy Spotlight Effects on Critical Time-Sensitive Diseases: Nationwide Retrospective Cohort Study on Taiwan's Hospital Emergency Capability Categorization Policy (e54651)
Chih-Yuan Lin, Chih-Ching Liu, Yu-Tung Huang, Yue-Chune Lee. . . . 798

XSL•FO
RenderX

Streamlining Ophthalmic Documentation With Anonymized, Fine-Tuned Language Models: Feasibility Study (e72894)
Sebastian Arens, Quang Ngo, Anna Richling, Lucas Stürzbecher, Daniel Böhringer, Thomas Reinhard, Felix Heilmeyer. 813

Current Status and Future Directions of Ferroptosis Research in Breast Cancer: Bibliometric Analysis (e66286)
Jia-Yuan Luo, Yu-Long Deng, Shang-Yi Lu, Si-Yan Chen, Rong-Quan He, Di-Yuan Qin, Bang-Teng Chi, Gang Chen, Xia Yang, Wei Peng. . . 8 2 4

Evolution of Artificial Intelligence in Medical Education From 2000 to 2024: Bibliometric Analysis (e63775)
Rui Li, Tong Wu. 857

Enhancing Recruitment of Adolescents Aged 16-18 Years in a Web-Based Peer Network Study Through Financial Reimbursements: Randomized Controlled Trial (e70223)
Sarah Eddy, Rachel Sacks-Davis, Michelle Raggatt, Cassandra Wright, Paul Dietze, Margaret Hellard, Jane Hocking, Megan Lim. 869

Increasing Participation and Completion Rates in Questionnaire Surveys of Primary Care Patients: Cluster-Randomized Study (e67981)
Paul Sebo, Benoit Tudrej, Augustin Bernard, Bruno Delaunay, Alexandra Dupuy, Claire Malavergne, Hubert Maisonneuve. 884

Association of Modifiable Lifestyle and Metabolic Factors With the Risk of Developing Sepsis: 2-Sample Mendelian Randomized Study (e72244)
Haifeng Lv, Jing Liu, Yelin Cao, Weina Fan, Guojie Shen, Feifei Wang, Qingqing Ye, Xiaoliang Wu, Kaijin Xu. 893

Correlation Between Objective Habit Metrics and Objective Medication Adherence: Retrospective Study of 15,818 Participants From Clinical Studies (e63987)
Antoine Pironet, L Phillips, Bernard Vrijens. 907

Impact of Fortified Malt-Based Food on Immunity Outcomes in School Children in India: Cluster Randomized Controlled Trial (e54189)
Anuradha Khadilkar, Vinay Rawat, Jaladhi Bhatt, Devyani Chaturvedi, Vivek Garg, Pankaj Verma. 916

A Proposed Taxonomy to Holistically Classify Employee Mental Health Programs: Qualitative Taxonomy Development Study (e67752)
Benedict Sevov, Robin Huettemann, Maximillian Zinner, Sven Meister, Leonard Fehring. 934

Perceptions and Experiences of Caregiver-Employees, Employers, and Health Care Professionals With Caregiver-Friendly Workplace Policy in Hong Kong: Thematic Analysis (e58528)
Maggie Lee, Eng-kiong Yeoh, Eliza Wong, Xue Bai, Nelson Yeung, Catherine French, Henock Taddese. 953

Understanding the Mental and Physical Burdens of Physicians and Identifying Support Interventions in Bangladesh: Qualitative Study (e76934)
Rahat Rony, Shams Aalok, Lamia Tisha, Marzan Mahatab, Nova Ahmed. 965

Compliance With Bowel Preparation and Its Influencing Factors in Patients Undergoing Colonoscopy: Cross-Sectional Study (e77189)
Huan Jiang, Chuanhui Li, Bing Hu, Yi Mou. 981

The Color of Drinking Survey Questionnaire for Measuring the Secondhand Impacts of High-Risk Drinking in College Settings: Validation Study (e64720)
Agustina Marconi, Reonda Washington, Amanda Jovaag, Courtney Blomme, Ashley Knobeloch, Vilma Irazola, Carolina Muros Cortés, Laura Gutierrez, Natalia Elorriaga. 992

XSL•FO
RenderX

Evaluating the Effect of the JUUL2 System With 5 Flavors on Cigarette Smoking and Tobacco Product Use Behaviors Among Adults Who Smoke Cigarettes: 6-Week Actual Use Study (e60620)
Nicholas Goldenson, Saul Shiffman, Mark Sembower, Arielle Selya, Steve Pype, Ryan Black. 1004

The Mediating Role of Depression in the Effect of Psychological Well-Being on the Self-Rated Health and Quality of Life of Older Adults: Cross-Sectional Study (e57731)
Meng Zhang, Ping Zhang, Qiu Xiang, Xiaopan Li, Suqin Xu. 1025

Association Between Prevention Focus and Sedentary Behavior in Older Adults: Cross-Sectional Study (e63280)
Jethro Suarez, Amber Blount, Kworweinski Lafontant, Joon-Hyuk Park, Rui Xie, Nichole Lighthall, Ladda Thiamwong. 1034

Associations Between Sleep Duration and Activity of Daily Living Disability Among Older Adults in China: Cross-Sectional Study (e65075)
Huimin Fan, Weijie Yu, Hongguo Rong, Xiaokun Geng. 1043

Multihealth Promotion Programs on Physical Health and Quality of Life in Older Adults: Quasi-Experimental Study (e65213)
Li-Yun Lee, Heng-Hsin Tung, George Liao, Su-Ju Liu, Zi-Yu Chen, Yea-Ru Yang. 1055

Predicting Age-Related Hearing Loss in Community-Dwelling Older Adults: Multicenter Retrospective Cohort Study (e81135)
Jing Li, Shuai Jin, Liu Sun, Jun-E Liu, Qiang Shen, Miao Shang, Hanting Wang, Yuanyuan Zhao. 1069

Investigating the Association Between Mean Arterial Pressure on 28-Day Mortality Risk in Patients With Sepsis: Retrospective Cohort Study Based on the MIMIC-IV Database (e63291)
Qimin Chen, Wei Li, Ying Wang, Xianjun Chen, Dehua He, Ming Liu, Jia Yuan, Chuan Xiao, Qing Li, Lu Chen, Feng Shen. 1083

French Versions of 2 English Questionnaires on Problematic Digital Use Assessed by Adolescents and Their Parents: Cross-Cultural Linguistic Translation and Adaptation Study (e55685)
Islam El Boudi, Mathilde Riant, Alexandre Bellier, Nicolas Vuillerme. 1095

Prevalence and Correlates of Clinically Elevated Depressive Symptoms in a Nationwide Sample of Transgender, Nonbinary, and Gender Diverse Young Adults in the United States: Cross-Sectional Survey Study (e66630)
Sari Reisner, Yuxin Liu, Regina Tham, Kaiden Kane, S Cole, Elizabeth Boskey, Sabra Katz-Wise, Alex Keuroghlian, Rena Xu. 1103

Citizen Worry and Adherence in Response to Government Restrictions in Switzerland During the COVID-19 Pandemic: Repeated Cross-Sectional Online Surveys (e55636)
Vanessa Kraege, Céline Dumans-Louis, Céline Maglieri, Séverine Bochatay, Marie-Anne Durand, Antoine Garnier, Kevin Selby, Christian von Plessen. 1115

French Versions of 4 English Questionnaires on Problematic Smartphone Use: Cross-Cultural Linguistic Translation and Adaptation Study (e53958)
Islam El Boudi, Mathilde Riant, Alexandre Bellier, Nicolas Vuillerme. 1126

Subtyping Service Receipt in Personality Disorder Services in South London: Observational Validation Study Using Latent Profile Analysis (e55348)
Jack Steadman, Rob Saunders, Mark Freestone, Robert Stewart. 1141

Detection of Chronic Cognitive-Motor Deficits in Adults With a History of Concussion Using Computerized Eye-Hand Coordination Tasks: Preliminary Experimental Design Study (e70867)
Qin Zhu, Shaochen Huang. 1160

XSL•FO
RenderX

Research Letters

Impact Factor Trends of Top Obstetrics and Gynecology Journals During COVID-19 (e70554)
Minhazur Sarker, Emily Yang, Ukachi Emeruwa, Timothy Wen. 1173

Population Estimates of Self-Reported Depression and Anxiety in the US From a National Survey: Cross-Sectional Survey Study (e70626)
Yuvraj Pathak, Elvira Makk-Frid. 1178

XSL•FO
RenderX

Corrigenda and Addenda

Correction: Digital Health for Australia: Bridging the Rural, Regional, and Remote Health Gap

Shakeel Mahmood[1], PhD; M Mamun Huda[1], PhD; Kedir Yimam Ahmed[1], PhD; Thapa Subash[1], PhD; Feleke Hailemichael Astawesegn[1], PhD; Anayochukwu Edward Anyasodor[1], PhD; Mohammad Ali Moni[1], PhD; Muhammad J A Shiddiky[1], PhD; Utpal K Mondal[1], MPH; Setognal Birara Aychiluhm[1], MPH; Santosh Giri[1], MPH; Allen G Ross[1], PhD

Rural Health Research Institute, Charles Sturt University, Orange, Australia

Corresponding Author:
Shakeel Mahmood, PhD
Rural Health Research Institute
Charles Sturt University
346 Leeds Parade
Orange, 2800
Australia
Phone: 61 0416689348
Email: shmahmood@csu.edu.au

Related Article:

Correction of: https://www.i-jmr.org/2025/1/e67460

Abstract

(Interact J Med Res 2025;14:e89755) doi:10.2196/89755

KEYWORDS

In “Digital Health for Australia: Bridging the Rural, Regional, and Remote Health Gap” [1], the authors noted one error.

The name of the final author in the authorship list originally read as follows:

Allen Ross

The final author’s name now reads:

Allen G Ross

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. Mahmood S, Huda MM, Yimam Ahmed K, Subash T, Astawesegn FH, Anyasodor AE, et al. Digital health for Australia: bridging the rural, regional, and remote health gap. Interact J Med Res 2025 Dec 02;14:e67460 [FREE Full text] [doi: 10.2196/67460] [Medline: 41329946]

XSL•FO
RenderX

Submitted 17.12.25; this is a non–peer-reviewed article;accepted 17.12.25; published 29.12.25.

Please cite as:
Mahmood S, Huda MM, Ahmed KY, Subash T, Astawesegn FH, Anyasodor AE, Moni MA, Shiddiky MJA, Mondal UK, Aychiluhm SB, Giri S, Ross AG
Correction: Digital Health for Australia: Bridging the Rural, Regional, and Remote Health Gap
Interact J Med Res 2025;14:e89755
URL: https://www.i-jmr.org/2025/1/e89755
doi:10.2196/89755
PMID:

©Shakeel Mahmood, M Mamun Huda, Kedir Yimam Ahmed, Thapa Subash, Feleke Hailemichael Astawesegn, Anayochukwu Edward Anyasodor, Mohammad Ali Moni, Muhammad J A Shiddiky, Utpal K Mondal, Setognal Birara Aychiluhm, Santosh Giri, Allen G Ross. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 29.12.2025. 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 Impact of Online Labor Platforms on Workforce Management in Health Care

Maryam Ahmadi Shad[1], MSc; Michael Simon[1], Prof. Dr; Florian Liberatore[2], Dr.

[1]Department of Public Health, Institute of Nursing Science, University of Basel, Basel, Switzerland

[2]Winterthur Institute of Health Economics, Zurich University of Applied Sciences, Gertrudstrasse 8, Winterthur, Switzerland

Corresponding Author:
Florian Liberatore, Dr.
Winterthur Institute of Health Economics, Zurich University of Applied Sciences, Gertrudstrasse 8, Winterthur, Switzerland

Abstract

Online labor platforms (OLPs) have the potential to change how the workforce is allocated and managed in health care. The contracting, coordination, and communication of bookings and work assignments happen on these platforms in near real-time with no delay and without any human interactions. This perspective paper describes the worldwide trend toward OLPs in health care, gives an overview of the functioning of these platforms, and discusses the prospects and challenges for health care management. As a real-world case, the platform logic, growth and traffic of a Swiss OLP designed for temporary nurse deployment are presented. OLPs facilitate managing different work arrangements (float pools and temporary work) through (1) offering health care staff flexible work options, which in turn lowers the dropout rates of health care professionals; and (2) effectively managing internal staffing allowing human resource sharing within and across health care organizations. For health care management research, OLPs yield data that can be used to analyze the characteristics, use, and dynamics of flexible work arrangements and temporary work in health care.

(Interact J Med Res 2025;14:e68546) doi:10.2196/68546

KEYWORDS

online labor platform; temporary work; workforce management; gig work; health care; personnel staffing and scheduling

Online Labor Platforms (OLPs): The Emerging Face of Temporary Employment in Health Care

For decades, health care organizations have relied on agencies to deploy temporary workers to fill their short-term vacancies and match workforce resources with care needs [1]. The use of temporary staff in hospitals is widespread globally. In the United States, about 75% of hospitals frequently employ temporary nurses [2]. Similarly, in the United Kingdom, the care hours provided by nonpermanent nurses increased more than doubled between 2012 and 2015 [3]. In Switzerland, though precise data are lacking, the use of temporary health care workers has risen since 2010. By 2022, temporary workers accounted for 3.5%-5.3% of health care staff, with agency nurses and care staff being the largest group [4]. With the emergence of OLPs, flexible work arrangements have undergone intense growth, and OLPs are increasingly replacing human match makers in temporary agencies for the intermediation between their temporary workers and their institutional clients.

In the United States, platforms such as Nomad, Medely, SnapCare, and CareRev have embraced the concept of temporary work by using web-based tools to connect hospitals with doctors and nurses for flexible, short-term assignments. Similarly, in Switzerland, web-based platforms like Careanesth, Coople, and Adecco serve as prominent avenues for the recruitment of temporary health care professionals.

Figure 1 shows the rising trend of annual shift bookings on Careanesth from 2016 to 2022. This trend has been largely neglected by health care managers as well as health services management research although OLPs could facilitate the efficient and effective use of temporary work and bring up new opportunities, challenges, and management issues for health care organizations with respect to processes and outcomes of recruitment, matching, administration, and evaluation of these arrangements.

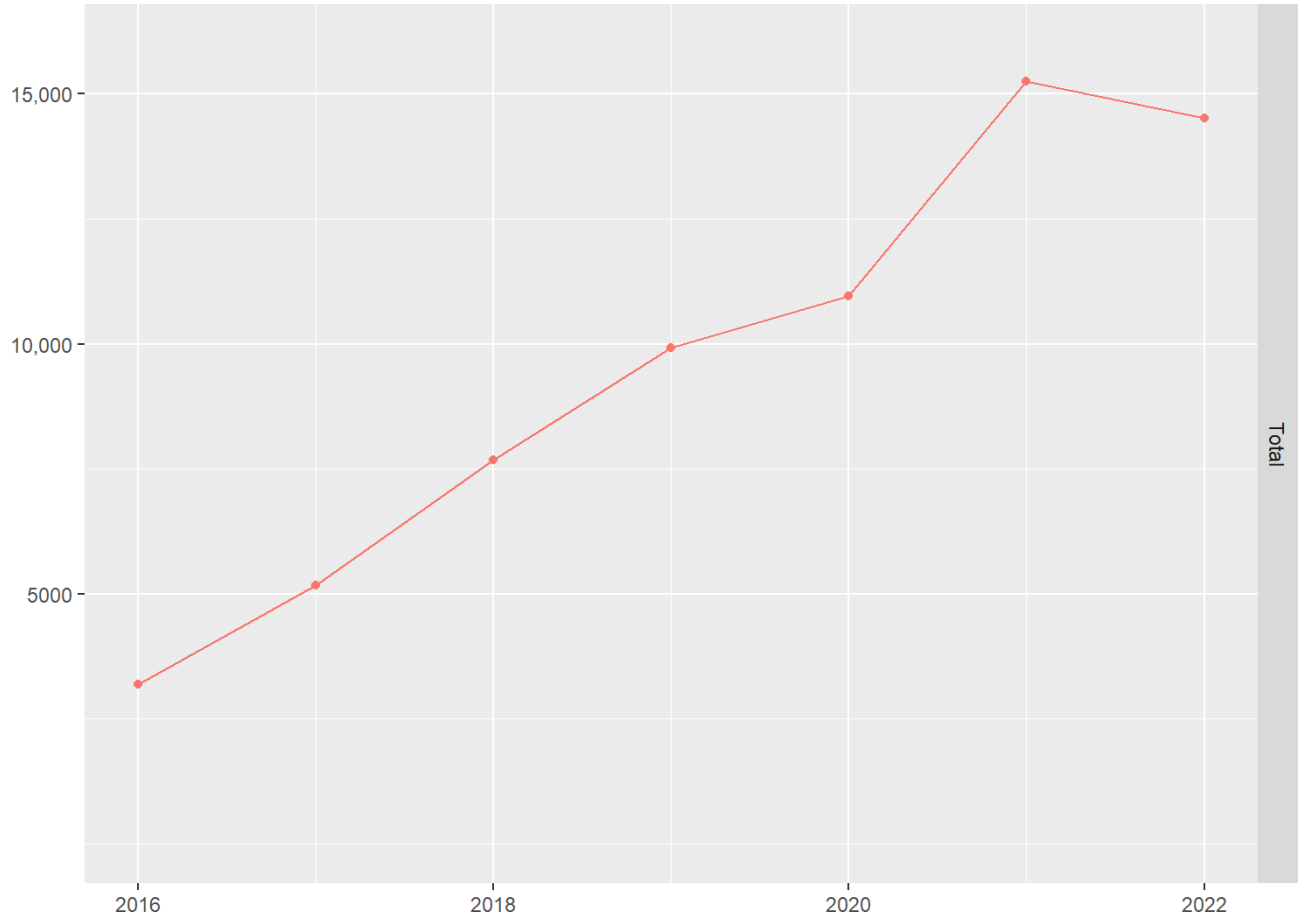
OLPs leverage information and communication technology (ICT) to make a continuous, real-time matching of staff availability with the demands [5]. Individuals can accept or decline work contracts as they choose, without obligations. OLPs vary widely in functionality: some of them allow individuals to transact, while others match between organizations and individuals. Some of them support and facilitate web-based work and others coordinate onsite or offline tasks at client locations, including in health care settings [6]. In health care, OLPs may be implemented for internal health professional float pools in a single organization, float pools across organizations but also for the booking of external temporary staff.

XSL•FO
RenderX

While there are substantial literature exploring temporary work via OLPs in various sectors, there remains a notable limitation of evidence specific to health care. This gap is critical because the nature of platform work in health care diverges significantly from other industries as mentioned. The onsite provision of services introduces unique challenges and complexities that are not as common in other sectors. Moreover, in health care, platform workers are often highly qualified professionals like doctors and nurses, unlike the less specialized workforce found in other sectors. Managing this specialized workforce requires advanced strategies, including strategic shift planning, rapid response to fluctuating demands, and coordination of diverse teams comprising various health care professionals. These complexities underscore the need for robust and adaptive workforce management strategies tailored to the unique demands of the health care sector. The evidence around integrating ICTs into the recruitment of health care professionals shows advantages such as cost-effectiveness, wider reach, and time-saving compared to traditional methods of recruiting health care professionals [7]. Although ICTs could have positive effects, their implementation may also provoke concern. The growth of digital labor platforms presents regulatory challenges, particularly in cross-border contexts [8]. The systematic review by Gagnon et al [9] provides a broader perspective, emphasizing the impact of ICTs on workforce sustainability. They can support the recruitment and retention of health care professionals by improving job satisfaction, streamlining communication, and reducing administrative burdens. They have the potential to transform workforce management in health care, from operational scheduling to strategic workforce planning, addressing both immediate and long-term needs [9]. Despite its relevance, there is limited research on this topic and considering that this is a rising phenomenon, the existing evidence is relatively dated. Moreover, the existing studies primarily address the broader integration of ICTs into workforce management in health care, often overlooking the specific and unique application of OLPs for deploying temporary professionals. These studies predominantly focus on outcomes, leaving many aspects of the underlying dynamics of OLPs unexplored. Therefore, there is a critical need to understand the dynamics of OLPs and how the integration of them affects various aspects of workforce management. This includes evaluating the reliability of current data, identifying gaps, and understanding the potential benefits that can be derived from more comprehensive and accurate datasets.

Our perspectives in this paper describe the worldwide trend toward OLPs in health care, gives an overview of the functioning of these platforms, and provides an in-depth discussion on the impact of OLPs on health care workforce management, focusing on staffing practices, demand responsiveness, and team dynamics. By providing insights into these areas, we aim to help improve management practices and enhance the efficiency and effectiveness of health care delivery, ultimately leading to high-quality patient care.

Figure 1. Growth of number of annual single shift bookings on a Swiss online labor platform.



XSL•FO
RenderX

Ethical Considerations

The empirical results provided in this paper have been the subject of an ethical consideration by the Cantonal Ethics Committee of the Canton of Zurich within the Crowis project. The study does not fall under the Human Research Act and an exemption of an ethical review was received (BASEC-Nr. Req-2022-00708). The use of the per-diem nurse data of the platform was based on informed consent with an opt-out system. All employees of the temporary agency were informed by a letter of information about the study and could dissent the use of their individual data. The employee IDs as identifiers of the entries in the platform data were replaced by an alphanumeric code by the temporary agency to guarantee confidentiality. Platform data were stored in a server protected by a password and accessible only to the members of the research team. No compensation was paid to the temporary nurses for data provision.

Operational Dynamics of OLPs and the Evolution of Workforce Management Practices

OLPs impact the whole process within the triangle of temporary agencies, temporary workers, and organizations booking services which can be divided into five steps: (1) enrollment in the platform; (2) offerings; (3) matching mechanisms; (4) administration; and (5) evaluation. The impact of OLPs on these steps is described in detail in the following.

Step 1: Enrollment

It typically refers to the process through which temporary workers register or sign up to become part of the platform's pool of available temporary workers. The registration requires a health care professional's basic personal information, professional credentials, and work experience. After the verification of the information and the certificate check by the platform, temporary workers may be required to review and sign agreements or contracts outlining the terms of their engagement with the platform, including payment terms, confidentiality agreements, and code of conduct. These profiles serve as a reference for health care facilities when they are looking to deploy temporary staff.

On the other side, organizations using OLPs need to provide key information to ensure smooth operations and compliance with platform standards. This includes clear details about shift offerings, payment terms such as hourly rates, payment methods, and timelines, along with guidance on administrative procedures. Additionally, organizations must share information about relevant labor laws and regulations that apply to health care workers, including their rights, protections, and available support mechanisms [6]. By transitioning from traditional agency-based enrollment methods to web-based platforms, temporary workers benefit from a more efficient, transparent, and accessible process for joining the health care workforce. It enhances communication and empowers health care professionals to manage and update their profiles effectively in real-time, reflecting changes in their professional status or preferences. Referring to the exemplary use case of the Swiss OLP, in 2022, 78 hospitals, 33 long-term facilities, 30 home-care organizations, and 11 other health care institutions were clients of the platform. On the platform, float pool as well as per-diem nurses who are employed by the platform provider can be booked for single shift assignments. A total of 283 per-diem nurses and 227 float-pool nurses were enrolled on the platform. Per-diem nurses can be booked from all institutions which are clients of the platform whereas float-pool nurses can be booked exclusively by each institution within their defined float pool, administered by the platform provider (Dettling M, Liberatore F. Unpublished data, April 2025).

Step 2: Offerings

Temporary workers and organizations can use a web browser or smartphone app to manage shift availability and vacancies (Figure 2). Health care professionals post the days and specific shifts they are available to work (eg, Monday, morning shift), while organizations post vacant shifts, specifying the ward type, required qualifications, and shift details (Figure 3).

Transitioning from the agencies to web-based platforms, health care professionals and hiring organizations have greater accountability and autonomy in leading transactions. Furthermore, OLPs provide the flexibility for hiring organizations to collect a group of "preferred" workers to prioritize them for job offerings. This differs from temporary agencies, where hiring organizations depend on human intermediaries in the agency to request temporary workers. This new model empowers both parties by eliminating the need for agency middlemen and streamlining the hiring process.

XSL•FO
RenderX

Figure 2. Managing shift offerings via the web browser or smartphone app.

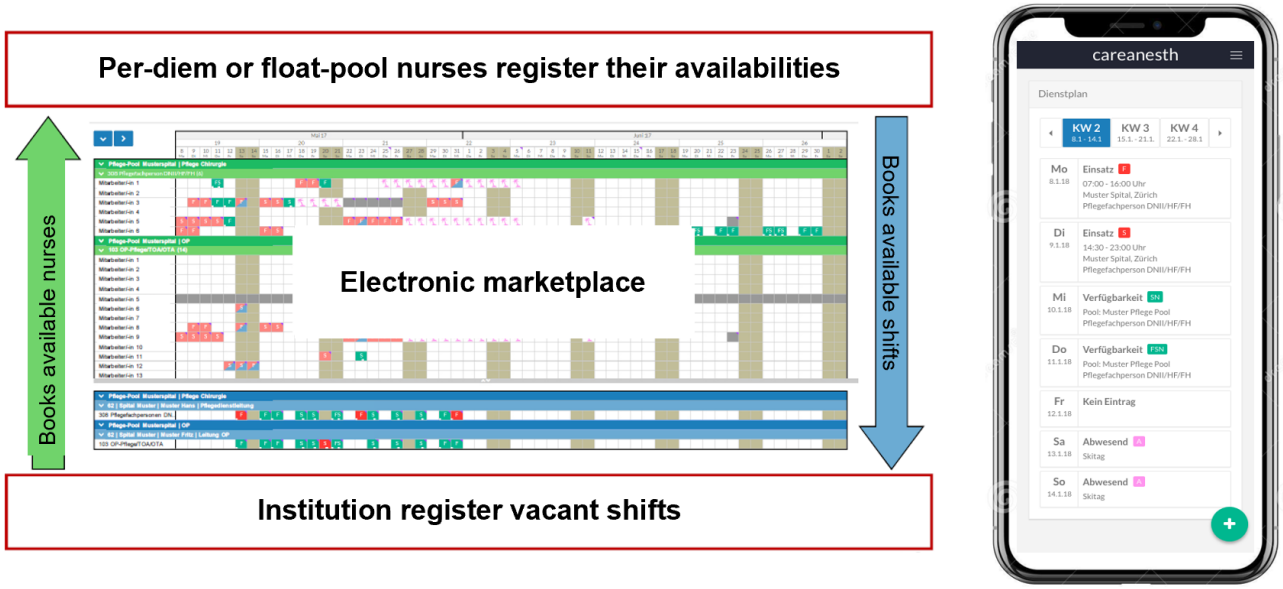
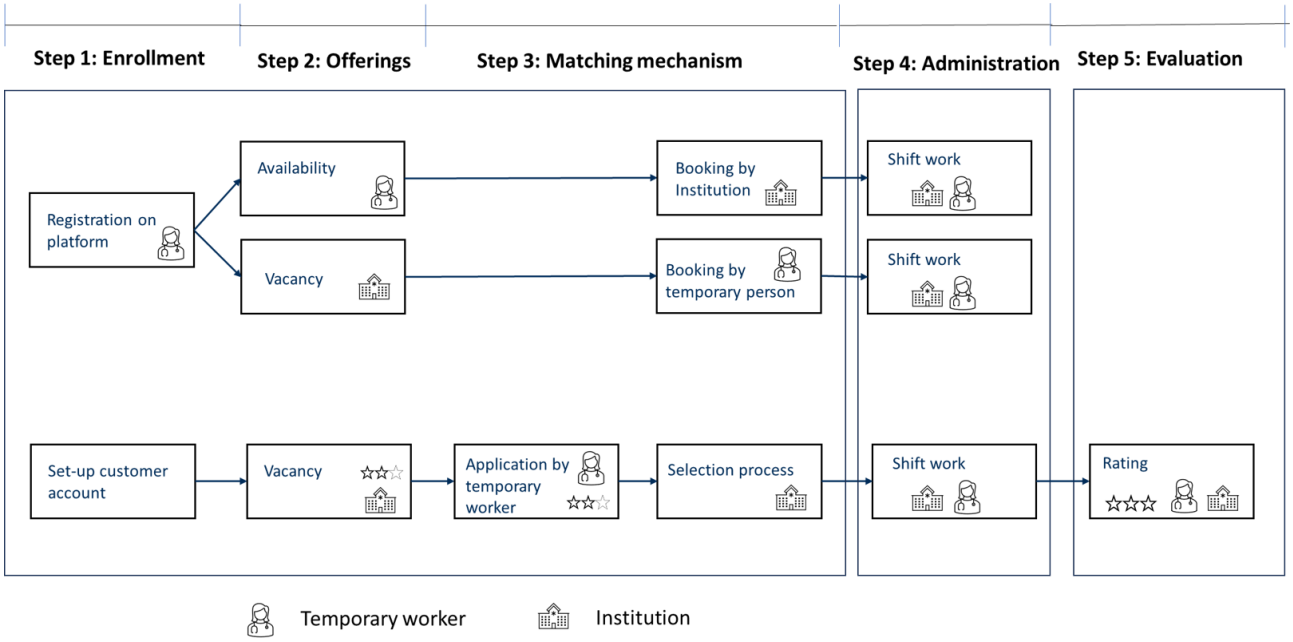


Figure 3. Overview about steps and general process on online labor platforms in health care settings.



Step 3: Matching Process

A match on a platform can either result from an organization booking a shift availability or from a temporary worker taking on a vacant shift offer. Both lead to a binding agreement regarding the shift assignment. However, in some OLPs, the initiation process stems solely from the organization's end, attracting a diverse pool of temporary job seekers. Here, applicants submit their applications, entering a selection process where the organization evaluates each individual against their criteria. This approach allows the organization to handpick the most suitable applicants. Matching mechanisms can follow the first come–first serve rule, which means that the first organization or temporary worker books the availability or vacant shift offer, and there is a binding agreement on the work. The exemplary case of the Swiss OLP uses this mechanism. In the booking process, the institutions see which type of nurse (registered nurse; vocational trained nurse assistant [Fachangestellte or Fachangestellter Gesundheit]; and intensive care nurse specialist [Intensivpflege]) is available for which shifts and have access to the individual curriculum vitae of the nurses to inform about qualifications and experience. Another possibility is that the organizations or temporary workers can apply for the offer and then after an evaluation by the contracting partner the contract is negotiated. (Figure 3). The web-based platform of Coople is an example that uses this mechanism. Alongside the bookings by hospitals and clients, some platforms use algorithms with and without artificial intelligence to analyze the data collected from both sides—organizations and professionals. These algorithms use various criteria and techniques to match the right health care worker with the right offer. Criteria may include skills, experience, location, availability, preferences, and historical performance data.

XSL•FO
RenderX

Platforms as managed service providers streamline the process of booking by allowing hospitals to access multiple temporary agencies through a single platform. This centralized approach simplifies staffing management and enhances efficiency for health care organizations. Hospitals can easily compare available temporary staff from various agencies, view their credentials, and make bookings according to their specific needs.

On OLPs, prebooking and prioritization rules are typically established to optimize the matching process between health care workers and organizations while ensuring efficient allocation of resources. These regulations may vary depending on the specific platform. These rules consider factors such as urgency of need (prioritize matches for immediate or time-sensitive vacant positions), availability (prioritize matches for only available workers), specialization, ratings, geographical proximity, repeated bookings in the past, premium services, and dynamic pricing (pricing algorithms may take into account factors such as demand, supply, and worker ratings to determine pricing for prebooked assignments, incentivizing workers to accept priority assignments) [7].

While OLPs primarily rely on ICT for matchmaking, traditional temporary agencies use "human matchmakers," such as agency representatives, to connect workers with short-term job opportunities. However, this does not mean temporary agencies avoid using technology. They use ICT to collect and manage data on their contingent workforce, including qualifications, work experience, and current assignments. Human match makers then use this information to assign workers to clients. In contrast, OLPs function without human intermediaries, offering a web-based marketplace where labor supply and demand directly meet [5].

Step 4: Administration

The contracting, coordination and communication of a work assignment happens in near real-time with no delay. The platform usually requests a fee for the service from the platform provider like fees for a personnel agency for the placement service or licensing fees for the use of the platform as health care organizations.

When comparing the contracting processes between traditional temporary agencies and OLPs, distinctions emerge, especially in how they engage with temporary workers. There exist 2 primary models within OLPs: co-employment platforms and freelance platforms, each with their distinct characteristics. Co-employment platforms establish an employment relationship with their workers, subjecting them to collective labor agreements either within the platform firm or with client organizations. Consequently, these platforms typically offer a fixed hourly pay rate to their temporary workforce, with potential surcharges during periods of labor scarcity. Conversely, freelance platforms engage solo self-employed workers who operate as independent contractors, allowing them to set their own hourly rates. The negotiation of pay rates can vary, with either the freelance worker or the client organization proposing rates depending on the platform's technical setup. At the same time, to avoid "races to the bottom," some of the freelance platforms' digital interfaces do not allow workers and hiring organizations to set hourly pay rates below a minimum level [5].

Additionally, automatic contracting as a key feature of OLPs facilitates the automatic generation and execution of contracts between the parties involved in a booking or transaction. This process is seamless and transparent, with the platform creating a standardized contract template based on the agreed-upon terms and relevant policies. These contracts typically outline key details such as service expectations, payment terms, duration, and other conditions. Once created, the contract is automatically sent to both parties for review and acceptance. This automated approach simplifies transactions, reduces administrative workload, and ensures clear communication and accountability for both parties.

In the use case of the Swiss OLP, the booking of a shift availability by an institution or the acceptance of a vacant shift by a float-pool nurse leads to a binding agreement between the institution and the temporary nurse. The matching (an assignment being offered and accepted) may occur as late as 24 hours before the start of the shift in question. The contracting, coordination and communication of a work assignment happens over the platform in near real-time with no delay.

Step 5: Evaluation

Some OLPs feature advanced ratings and reputation systems that foster trust and transparency between users. The Quality Rating 360 empowers both parties to provide comprehensive feedback on performance, behavior, and competencies, ensuring a fair and reliable marketplace. Temporary workers can rate organizations on factors like work environment, communication, and overall satisfaction, helping guide fellow professionals in making informed choices. Conversely, organizations can rate temporary workers based on professionalism, performance, and reliability, contributing to a dynamic feedback loop that enhances platform quality and accountability. Unlike traditional agencies, this feature grants the users valuable insights into the reputation and performance of various entities, enhancing transparency and informed decision-making. In the use case of the Swiss OLP, no rating system has been established to date but is planned.

The Prospects and Challenges of OLPs for Health Care Management

Impact of Using OLPs

The use of OLPs for temporary deployments has a significant impact on organizational management endeavors related to shift planning, the availability of new work arrangements, patient flow management, and the opportunities to share resources within and across organizations. However, the impact is associated not only with prospects but also with new challenges for health care managers.

Managing Flexible Workforce in Shift Planning

In terms of workforce management, there are both benefits and challenges for hospitals whose workforce includes temporary staff. On the one hand, the web-based booking mechanisms, facilitate the staffing of shifts according to demand and make

XSL•FO
RenderX

the covering of short-term absences more efficient (eg, by requiring fewer phone calls and emails) [8]. On the other hand, the operations on these platforms bring new administrative burdens for managers. They must draft and publish vacant shifts, and select temporary health care workers according to their shift demands, and availabilities of temporary staff [5]. Competencies and work experiences of temporary health care workers must be assessed based on the provided nurse-related information on the platform without personal contact or assessment [9]. Further, on OLPs, availabilities for unpopular shifts (eg, night shifts or shifts on public holidays) may be much less. Because OLPs act as a labor market, shift managers must adopt selective booking behaviors. This means that they do not book temporary workers for popular shifts and instead post vacancies for unpopular shifts, and temporary workers are forced to adapt their operational availability to the demands of the hospital in order to work there. Moreover, the monitoring of compliance with maximum working hours regulations and rest periods is challenging for health care institutions because the temporary health care workers may work in various institutions based on multiple work contracts with agencies and health care institutions in parallel.

Offering New Work Arrangements for Health Care Professionals Using OLPs

The digital platform economy provides opportunities for new and diverse work arrangements. It potentially enables mixed contracts where individuals can engage in part-time, freelance, or fixed work simultaneously, reflecting the evolving nature of work in this digital era. The web-based booking systems save temporary workers from being permanently on call. Either they offer their shift availabilities which are most suitable to their individual preferences and obligations [10] or they consider overtaking vacant shifts offered by organizations that pop up on the platform. The decisions can be made without any obligations. This feature of self-scheduling in new work arrangements has been associated with higher job satisfaction and retention. However, nurse managers must cope with greater complexity due to the mixture of various temporary work arrangements in the shift planning. The lack of organizational boundaries in the work arrangements poses the challenge of assessing the competencies and experiences of temporary nurses without personal contact or assessment [9]. Health care managers must find solutions to overcome the resistance to change and install working contexts in which ad hoc teams with a mixture of permanent and temporary staff can work together in an efficient and effective way.

While for some health care professionals, these working models offer new opportunities, others may prefer traditional working and staffing models. Platform work increases the potential for worker exploitation and the workers have to deal with challenges caused by platform-based vulnerabilities such as inequality and precarity; which refers to the short-term contingent nature of the work and lack of clear training and career promotion [11,12]. As platform workers are most likely to be considered self-employed, rules on data protection and privacy become challenging. There is very little information on how platforms protect data and the privacy of workers as well as organizations. Concerns have also been raised that platforms possess an inequitable advantage over conventional agencies, as platforms have fewer responsibilities for working conditions such as issues related to the data privacy and intellectual property rights of temporary workers [6], making it challenging for health care institutions to comply with these regulations using temporary health care workers from these platforms. Furthermore, issues such as wage discrimination can occur on OLPs during the matching process due to varying wage structures. Some platforms let organizations set hourly rates, potentially leading to inconsistencies. Others offer fixed rates, possibly undervaluing experienced temporary workers. Platforms with flexible rates based on factors like experience may unintentionally perpetuate disparities. Additionally, fluctuating rates based on demand-supply dynamics could result in unequal pay.

Sharing Resources Within and Across Organizations Using OLPs

OLPs are instrumental in promoting resource sharing within and across health care organizations. They facilitate sharing health care professionals across different organizations or units using 1 single tool and without the need for human matchmaking in each organization or unit. This approach aligns with the contemporary trend of using technology to bridge gaps and streamline interactions within and across organizations. Regarding sharing the workforce within and across organizations through OLPs, several challenges arise. Appropriate rules and mechanisms for prioritization, booking or prebooking, cost allocation, and activity tracking need to be defined to optimize resource use and enhance user experiences in the sharing economy paradigm. Prioritization in workforce management is essential for efficiently allocating resources and ensuring that the right people are in the appropriate roles at the correct time. Further, yield management and dynamic pricing approaches may be applied within the booking mechanisms of OLPs to foster the market mechanisms on these platforms resulting in a better allocation of the shared resources. Yield management as a dynamic pricing approach can be used to control capacities by adjusting prices based on real-time demand, supply, and market conditions. The sharing of health care professionals across organizations may be challenging because of obligatory minimum staffing regulations and its verification by nurse-patient ratios depending on the health care system setting.

Effects of OLP Application on Patient Flow Management

Another significant advantage offered by OLPs is their facilitation of patient flow management and staff rescheduling within health care organizations. These platforms provide real-time access to data regarding the availability of health care professionals. This enables facilities to promptly address staffing gaps and adjust patient flow plans and staffing levels, accordingly, ensuring appropriate staffing levels in unexpected surges in demand [5]. Furthermore, the retrospective data capabilities of OLPs facilitate proactive workforce planning for both fixed and temporary staffing level requirements. By examining the trends and patterns, health care facilities can potentially forecast future demands and make strategic decisions about resource allocation to effectively address patient needs.

XSL•FO
RenderX

This data-driven information enables organizations to align their staffing schedules more precisely with anticipated demand, thereby maximizing resource utilization and boosting the overall efficiency of patient flow management.

Propositions for Future Research on Online Health Care Labor Platforms

OLPs and their application for the recruiting and management of temporary health care workers offer several propositions for future research.

First, studies should calculate and compare changes in administrative costs (cost savings vs new administrative burdens) using OLPs for the application of temporary work arrangements in health care organizations. Second, future research should address the impact of different offering and matchmaking rules on OLPs on matching rates and labor costs to understand better how OLPs allow health care organizations to apply and rely on dynamic shift planning systems with a mix of permanent and flexible staff. Third, the consequences of the increasing use of short-term single-shift assignments arranged via OLPs on treatment processes, costs, and outcomes require more research. Fourth, the market dynamics on OLPs and the leverage of health care organizations altering the supply side on these platforms is an important area informing health care managers how to behave optimally on these OLPs. Fifth, OLPs yield data that can be used to analyze the characteristics and dynamics of temporary work contributing to a greater understanding of their behavior as well as the use of temporary nurses in organizations. Compared to findings from surveys of temporary agency managers and temporary health care workers yielding motivations and behavioral intentions for temporary work, data from platforms represent actual behavior and therefore offer a sound basis for discussion on the extent and patterns of temporary work in a health care system. More specifically, platform data includes the following information.

- Data about the organizations and temporary health care workers operating on the platform inform about the sociodemographics of the temporary health care workers and organizational characteristics.
- Shift availabilities show how many and which shifts temporary health care workers offer thereby revealing their work preferences. Shift offers by the organizations on the platforms inform about the volume and characteristics of unstaffed shifts.
- Matching data displays how often temporary health care workers are booked, in how many different organizations they work, and in which types of shifts they are deployed. Further, matching rates provide information about the effectiveness of the platforms matching temporary health care workers' availabilities and shift demands.

Overall, it can be concluded from the perspective paper, that health services management research should increase research activities in the field of OLPs in health care. The potential of allocating health care workers to their needs, but also according to the demands of the health care organization is a key topic for health care systems worldwide.

Acknowledgments

This research as a part of the CroWiS project, was supported by the Swiss National Science Foundation through the National Research Program "Digital Transformation" (NRP 77; grant 187433). The authors gratefully acknowledge their support.

Conflicts of Interest

None declared.

References

1. Alanazi FK, Lapkin S, Molloy L, Sim J. Healthcare-associated infections in adult intensive care units: A multisource study examining nurses' safety attitudes, quality of care, missed care, and nurse staffing. Intensive Crit Care Nurs 2023 Oct;78:103480. [doi: 10.1016/j.iccn.2023.103480] [Medline: 37379679]
2. May JH, Bazzoli GJ, Gerland AM. Hospitals' responses to nurse staffing shortages. Health Aff (Millwood) 2006 Jan;25(Suppl1):W316-W323. [doi: 10.1377/hlthaff.25.w316]
3. Senek M, Robertson S, Ryan T, King R, Wood E, Tod A. The association between care left undone and temporary nursing staff ratios in acute settings: a cross- sectional survey of registered nurses. BMC Health Serv Res 2020 Jul 10;20(1):637. [doi: 10.1186/s12913-020-05493-y] [Medline: 32650763]
4. Job placement and staff rental. Obsan. 2024. URL: https://ind.obsan.admin.ch/indicator/pflemo/arbeitsvermittlung-und-personalverleih [accessed 2024-09-16]
5. Meijerink J, Arets M. Online labor platforms versus temp agencies: what are the differences? SHR 2021 Oct 1;20(4):119-124. [doi: 10.1108/SHR-12-2020-0098]
6. Mandl I, Kilhoffer Z, Lenaerts K, De Groen WP, Eurofound. Employment and working conditions of selected types of platform work. Publications Office of the European Union. 2018. URL: https://data.europa.eu/doi/10.2806/929950 [accessed 2024-01-22]
7. Gautam KS. A summation of online recruiting practices for health care organizations. Health Care Manag (Frederick) 2005 Jul;24(3):257-267. [doi: 10.1097/00126450-200507000-00010]

XSL•FO
RenderX

8. Berg J, Cherry MA, Rani U. Digital labour platforms: a need for international regulation? Revista de Economía Laboral 2019 Dec 31;16(2):104-128. [doi: 10.21114/rel.2019.02.05]
9. Gagnon MP, Pollender H, Trépanier A, Duplàa E, Ly BA. Supporting health professionals through information and communication technologies: a systematic review of the effects of information and communication technologies on recruitment and retention. Telemed J E Health 2011 May;17(4):269-274. [doi: 10.1089/tmj.2010.0175] [Medline: 21476881]
10. Lien CY. Temp nurses go digital: Examining gig care in US nursing homes. Sociol Health Ill 2023 Mar;45(3):542-559. [doi: 10.1111/1467-9566.13600]
11. Bajwa U, Gastaldo D, Di Ruggiero E, Knorr L. The health of workers in the global gig economy. Global Health 2018 Dec 18;14(1):124. [doi: 10.1186/s12992-018-0444-8] [Medline: 30563539]
12. Marmo M, Sinopoli EA, Guo S. Worker exploitation in the Australian gig economy: emerging mechanisms of social control. Griffith Law Review 2022 Apr 3;31(2):171-192. [doi: 10.1080/10383441.2022.2076036]

Abbreviations

ICT: information and communication technology
OLP: online labor platform

Edited by TDA Cardoso; submitted 11.11.24; peer-reviewed by D Belias, S Balaguru; revised version received 27.02.25; accepted 27.02.25; published 02.05.25.

Please cite as:
Ahmadi Shad M, Simon M, Liberatore F
The Impact of Online Labor Platforms on Workforce Management in Health Care
Interact J Med Res 2025;14:e68546
URL: https://www.i-jmr.org/2025/1/e68546
doi:10.2196/68546

© Maryam Ahmadi Shad, Michael Simon, Florian Liberatore. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 2.5.2025. 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

Systemic Inflammation and Disruption of the Local Microenvironment Compromise Muscle Regeneration: Critical Pathogenesis of Autoimmune-Associated Sarcopenia

Yingjuan Zhang[1*], MSc; Qingqian Wu[1*], PhD; Yi Wang[2*], BSc; Qingyan Chen[3], BSc; Shuang Han[1], PhD; Bei Li[1], PhD; Qingwen Zhao[1], PhD; Qianzhuo Wang[1], BSc; Yule Wang[1], PhD; Yue Gao[1], BSc

[1]Department of Geriatrics, Zhejiang Key Laboratory of Traditional Chinese Medicine for the Prevention and Treatment of Senile Chronic Diseases, Affiliated Hangzhou First People's Hospital, School of Medicine, Westlake University, 261 Huansha Road, Shangcheng District, Hangzhou, China

[2]Yixing Center for Disease Control and Prevention, Yixing, Jiangsu Province, China

[3]Fourth Clinical School of Medicine, Zhejiang Chinese Medical University, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Yue Gao, BSc
Department of Geriatrics, Zhejiang Key Laboratory of Traditional Chinese Medicine for the Prevention and Treatment of Senile Chronic Diseases, Affiliated Hangzhou First People's Hospital, School of Medicine, Westlake University, 261 Huansha Road, Shangcheng District, Hangzhou, China

Abstract

Sarcopenia is defined by age-related reductions in muscle mass, strength, and physiological function, and it is especially prevalent among individuals with autoimmune diseases. Autoimmune disorders, characterized by immune dysregulation, cause systemic inflammation and damage to multiple tissues through unregulated immune activity. Research indicates that autoimmune diseases negatively impact skeletal muscle functions and may worsen the progression of sarcopenia. This viewpoint comprehensively discusses the pathogenesis and potential mechanism of sarcopenia in 3 autoimmune diseases: inflammatory bowel disease, rheumatoid arthritis, and type 1 diabetes mellitus. Mechanistically, chronic immune microenvironment alterations induce compartment-specific redistribution of leukocyte subsets and cytokine networks. These perturbations disrupt critical signaling pathways governing muscle protein synthesis, satellite cell activation, and mitochondrial bioenergetics, leading to impaired regeneration and accelerated sarcopenia progression. By delineating shared and distinct pathomechanisms across these models, this analysis reframes our understanding of immune-mediated muscle wasting. Beyond mechanistic insights, it establishes a translational framework for targeted therapies and highlights emerging research directions bridging immunology and age-related musculoskeletal decline.

(Interact J Med Res 2025;14:e64456) doi:10.2196/64456

KEYWORDS

sarcopenia; immune diseases; type 1 diabetes mellitus; rheumatoid arthritis; inflammatory bowel disease

Introduction

Sarcopenia is an age-related syndrome characterized by progressive and systemic loss of skeletal muscle mass and strength [1]. The decline of skeletal muscle function caused by sarcopenia has aroused widespread concern [2]. Maintaining skeletal muscle mass and integrity is essential for the proper functioning of the musculoskeletal system and for the effective absorption and storage of nutrients [3]. Sarcopenia has a significant impact on general health, with numerous important effects. It is closely related to the development of exercise-related diseases, such as falls, fractures, and disabilities [4]. Additionally, it is also associated with adverse outcomes such as cardiovascular diseases [5], metabolic diseases, and even death [6]. However, the exact cause of primary sarcopenia is not fully understood yet. Recent studies suggested that cellular aging [7], mitochondrial disorder [8], decreased physical activity [9], and excessive caloric intake could accelerate the progression of sarcopenia. Moreover, an imbalance of oxygen in the body may contribute to the development of sarcopenia as well [10].

Sarcopenia afflicts not only the older people but also the patients with immune diseases. Studies have shown that immune diseases such as type 1 diabetes mellitus (T1DM), rheumatoid arthritis (RA), inflammatory bowel disease (IBD), spinal arthritis, and systemic sclerosis are often associated with sarcopenia [11]. Autoimmune diseases combined with sarcopenia are closely related to adverse clinical outcomes, poor effects of treatment and prognosis, and low quality of life [12]. Current research efforts have concentrated on the high prevalence of sarcopenia among autoimmune disease patients and its associated adverse

XSL•FO
RenderX

prognosis. Studying the relationship between sarcopenia and autoimmune diseases is crucial, as is clarifying the incidence and pathogenesis of sarcopenia linked to immune disorders. Immune disorders regulate key cascades like JAK/STAT (Janus kinase/signal transducers and activators of transcription) by influencing immune cells, profoundly impacting muscle protein synthesis and metabolism. This is a crucial factor in the pathogenesis of autoimmune diseases accompanied by muscular atrophy. Previous studies have focused on the differences in performance between healthy and sick individuals through case-control analysis. However, our comprehensive and in-depth analysis shows that focusing on immune cells and the related factors released and exploring the pathogenesis of diseases based on this, to find effective targeted therapies, will become a new and promising direction for the study of autoimmune diseases and sarcopenia. The elucidation of these underlying mechanisms may pave the way for novel research avenues in understanding the pathophysiology of sarcopenia associated with autoimmune disorders, while simultaneously offering crucial mechanistic insights to inform the development of targeted therapeutic interventions with enhanced precision. This study aims to provide a theoretical foundation and new avenues for future research.

Sarcopenia

In 1989, Irwin Rosenberg first described sarcopenia as the loss of muscle mass that occurs with aging [13]. With the deepening of research, sarcopenia is generally recognized as a progressive and systemic skeletal muscle disease manifested by accelerated loss of muscle mass and function [14]. The European Working Group on Sarcopenia in Older People 2 updated the definition of sarcopenia to account for muscle mass, muscle strength, and motor function performance for the first time [13]. The European Working Group on Sarcopenia in Older People 2 suggests that sarcopenia may be present when muscle strength is insufficient, and can be diagnosed when low muscle quantity or quality exist simultaneously [13]. Moreover, when the patient has low muscle mass, low muscle strength, and weak motor function at the same time, severe sarcopenia can be diagnosed [15]. Applying Western diagnosis criteria to Asian people may not be adequate. For special considerations such as different body sizes, fat distribution, and physical activity from the Asians, the Asia Working Group for Sarcopenia proposed and revised an algorithm for diagnosing sarcopenia based on the Asian data [1,16] (Figure 1).

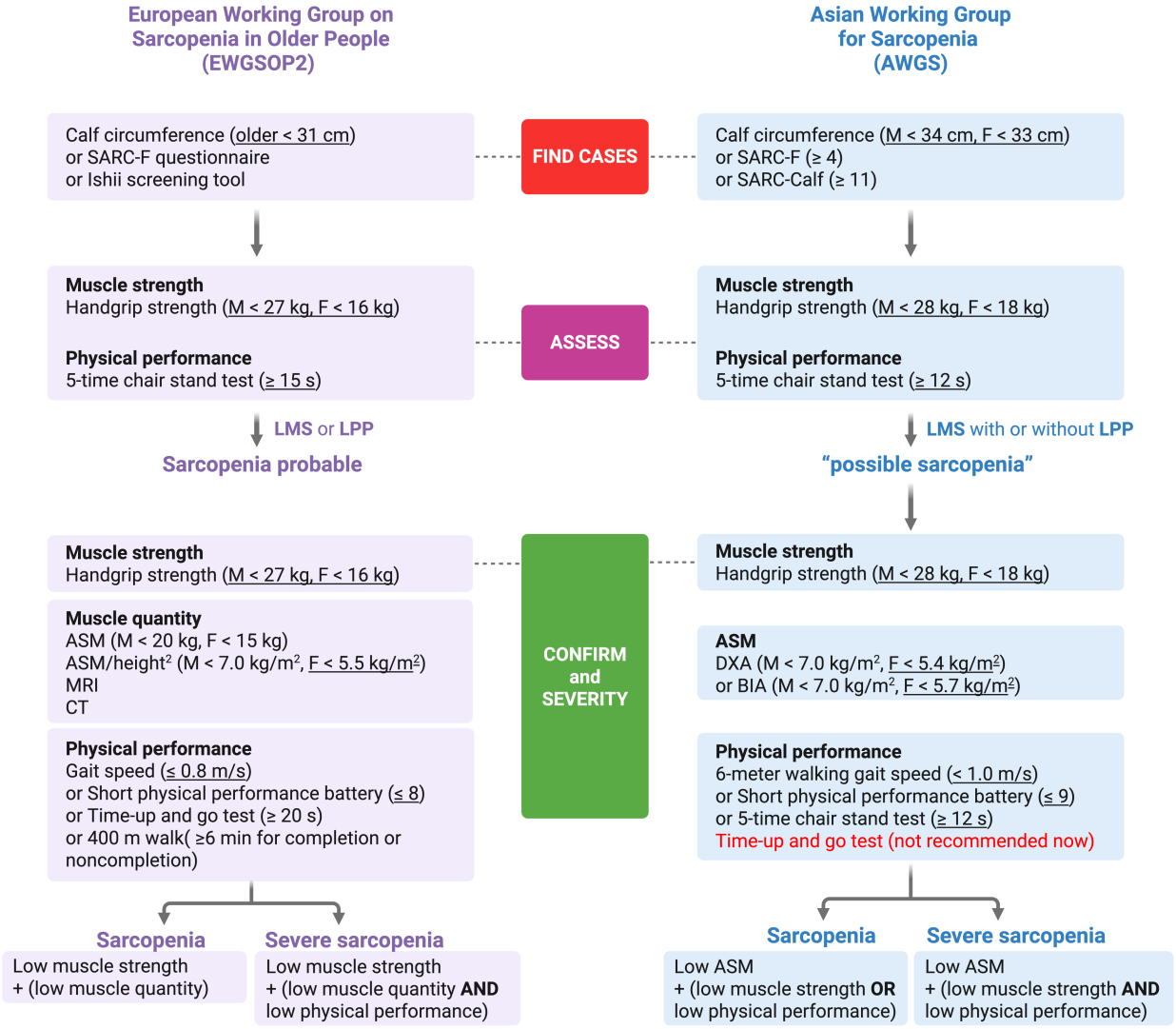
The prevalence of sarcopenia in older individuals within community health care settings has been reported to reach up to 29% [16]. Aging changes the homeostasis of skeletal muscle and disturbs the balance between anabolic and catabolic processes on the protein production pathway [16]. Sarcopenia is characterized by a decrease in the size and number of satellite cells and the atrophy of type 2 muscle fibers. Sarcopenia is distinct from malnutrition and cachexia, although each exhibits interrelated pathophysiology leading to different magnitudes of “wasting” and susceptibility to cardiovascular events (Table 1) [2].

Diagnosing sarcopenia requires a combined measurement of muscle mass, strength, and physical performance [17] (Figure 1). At present, imaging methods, such as computed tomography, dual-energy x-ray absorptiometry (DXA), and magnetic resonance imaging are used for muscle mass detection [18]. However, those instruments are expensive, time-consuming, and laborious. Besides, computed tomography detection has a certain level of radioactivity, which limits its application in clinical practice and population research [19]. DXA can distinguish adipose tissue from other soft tissues and measure the volume of tissues and organs accurately, so it is suitable for different populations and is considered the “gold standard” for measuring muscle mass [20]. Owing to the characteristics of fast, convenient, and low radioactivity, DXA is also frequently used for body composition detection. However, DXA has the limitations of not assessing muscle quality (ie, muscle fat infiltration), and DXA measurements may be influenced by the patient’s hydration status [21]. The machine requires a large space and is inconvenient to move away, making DXA not an ideal measurement for large-scale clinical trials of sarcopenia, bedside assessments, or community health screening work [18]. Bioelectrical impedance analysis (BIA) is another measurement for muscle mass detection. The principle underlying BIA is that different body compositions at the tissue-organ level exhibit varying electrical conductivity, with skeletal muscle demonstrating good conductivity and adipose tissue showing poor conductivity. This enables BIA to estimate the quantity of body fat and skeletal muscle mass [22]. Comparatively speaking, except inexpensive, BIA is a portable device that can easily perform bedside diagnostics as well. Therefore, BIA is frequently used in many types of studies. As BIA equations and cut-off values are population and device-specific, results can vary between devices and different populations [23]. Moreover, studies showed that the mass of hospitalized older patients was overestimated [24]. Nevertheless, the results of BIA would also be affected by the placement of electrodes, so it is necessary to combine grip strength and walking speed detection to evaluate the occurrence of sarcopenia [25].

Sarcopenia increases the risk of falls, fractures, and death. When combined with other diseases, it can hinder recovery after hospitalization, prolong prognosis recovery time, and increase the economic burden on patients [26]. Previous studies suggested that sarcopenia and osteoporosis interacted with each other. Compared with healthy people, individuals with sarcopenia have a heightened susceptibility to developing osteoporosis. Conversely, individuals experiencing severe osteoporosis often undergo substantial loss of skeletal muscle mass, thereby increasing their vulnerability to falls and fractures [27]. Sarcopenia is known to adversely affect the outcomes of patients with various forms of cancer. Apart from what is mentioned above, there exists an interaction between sarcopenia and immune diseases. The complication of sarcopenia greatly increases the incidence; mortality; and prognosis of diabetes, arthritis, and IBD.

XSL•FO
RenderX

Figure 1. Simple algorithms for Europeans and Asians to diagnose sarcopenia in clinical practice respectively. ASM: appendicular skeletal muscle mass; BIA: bioelectrical impedance analysis; CT: computed tomography; DXA: dual-energy x-ray absorptiometry; F: female; LMS: low muscle strength; LPP: low physical performance; M: male; MRI: magnetic resonance imaging; SARC-Calf: SARC-F plus calf circumference (score of ≥11 is termed as sarcopenia); SARC-F: strength, assistance in walking, rising from a chair, climbing stairs, and falls (score of ≥4 is termed as sarcopenia).



XSL•FO
RenderX

Table . Differences between sarcopenia, cachexia, and malnutrition.

	Sarcopenia	Cachexia	Malnutrition
Clinical features	Loss of muscle strength and muscle mass	Weight loss with loss of muscle mass	Weight loss
Functional impairment	Indicate severity	+++[a]	–[b]
Mechanism	Age-related, pathologic	Pathologic	Inadequate caloric intake, malabsorption
Inflammation	+/–[c]	+++	–
Anorexia	+/–	++[d]	+/–
Fat mass	+/–	Decreased	Decreased
Protein degradation	+/–	+++	+[e]

[a]+++: high, a higher degree or intensity of a feature or symptom in the disease.

[b]–: absence of the characteristic/symptom indicates the absence or absence of a characteristic, symptom, or mechanism.

[c]+/–: variable presence or inconsistency across conditions.

[d]++: moderate, the presence or intensity of a characteristic or symptom in the disease.

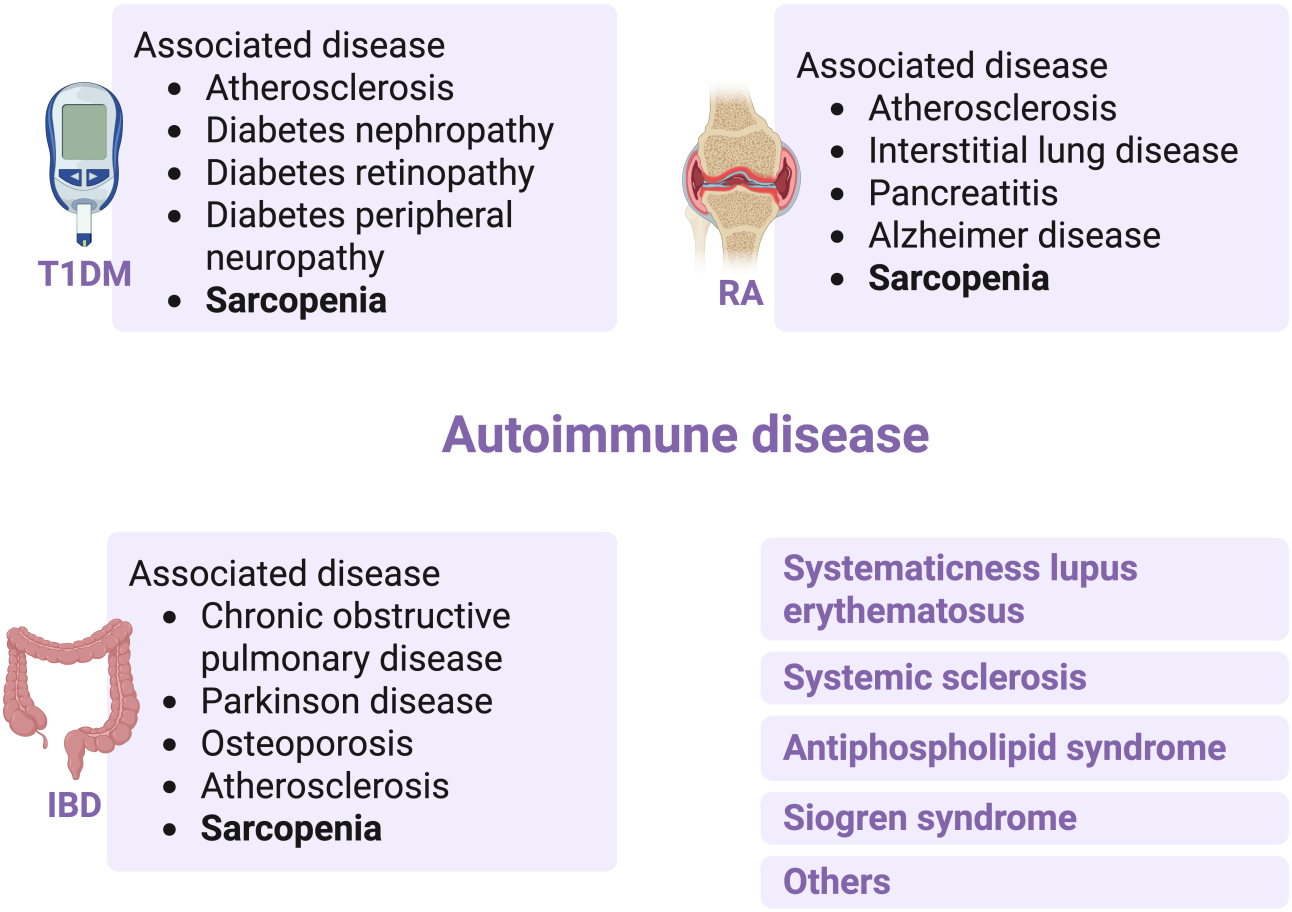
[e]+: low, a feature or symptom that is less present or intense in the disease.

Autoimmune Diseases

Autoimmune diseases are a series of related diseases characterized by dysregulation of the immune system, in which immune cells cannot distinguish self-antigens from non–self-antigens and give rise to activation of immune cells to attack autoantigens, leading to inflammation and multi-tissue damage [28]. T1DM, RA, and IBD are autoimmune diseases [29]. Moreover, some of the autoimmune diseases are organ-specific, such as primary biliary cirrhosis, and some reflect a variety of immunological dysfunctions involving multiple organ immune dysfunction such as systemic lupus erythematosus [30]. T1DM is dramatically increasing in many parts of the world [31], they affect millions of people and are prone to serious complications such as heart and kidney disease, stroke, and blindness [32] (Figure 2). RA is an autoimmune disease that damages joints and other tissues and organs such as the heart, kidneys, lungs, digestive system, eyes, skin, and nervous system [33]. IBD is a chronic, progressive immune-mediated inflammatory condition of the gastrointestinal tract [34]. The incidence of IBD has rapidly increased worldwide. Patients with IBD may suffer from malnutrition, characterized by an energy or nutrient imbalance, leading to sarcopenia, micronutrient deficiencies, overweight, obesity, and sarcopenic obesity [35]. Sarcopenia is a common complication of autoimmune diseases, and the primary cause is the loss and damage of skeletal muscle induced by immune dysfunction [36-38].

XSL•FO
RenderX

Figure 2. The associated disease of autoimmune diseases. IBD: inflammatory bowel disease; RA: rheumatoid arthritis; T1DM: type 1 diabetes mellitus.



Type 1 Diabetes Mellitus and Sarcopenia

T1DM results from the autoimmune dysfunction and death of β-cells, requiring lifelong insulin therapy [39]. This condition arises from a complex interaction between invading or resident macrophages and T cells. These immune cells release various substances, such as chemokines and cytokines, in the islet microenvironment that promote the apoptosis of beta cells. Simultaneously, β-cells can attract and activate immune cells to the islet area through their signaling, especially in response to stress, damage, or cell death [40]. The primary consequence of T1DM is hyperglycemia. Research indicates that the insulin-like growth factor (IGF) and IGF-binding protein (IGFBP) axis are crucial in maintaining normal glucose levels. Patients with T1DM showed a significant reduction in serum IGF-I levels. Additionally, mitophagy plays a crucial role in regulating the autoimmune response during the development of T1DM by preventing the accumulation of defective or dysfunctional mitochondria in pancreatic cells [41]. Muscle fibers can be damaged and undergo adverse remodeling due to the hyperglycemic and inflammatory microenvironment caused by T1DM, ultimately leading to sarcopenia [42]. A cross-sectional study showed that the prevalence of sarcopenia in participants aged 65 years or older with T1DM is 42.9%, which is significantly higher than that in people of the same age without sarcopenia [43]. Moreover, musculoskeletal system health would be affected if T1DM was poorly controlled in childhood and adolescence, and then sarcopenia would become the late complication of T1DM in adults [42].

Skeletal muscle is a primary driver organ that maintains whole-body glycemic balance [44]. Poor blood glucose control is the main risk factor for T1DM, which is complicated by sarcopenia [45]. IGFs, especially IGF1 and IGF2, can promote glucose metabolism, with their availability regulated by IGFBPs. In sum, IGFs are dysregulated both before and after the clinical diagnosis of T1DM and may serve as novel biomarkers to improve disease prediction [46]. Subsequently, Hata et al [47] discovered that a low serum IGF-1 level is associated with sarcopenia and low skeletal muscle mass in subjects with T1DM. Moreover, growth hormone could regulate cell survival and the musculoskeletal system function through growth hormone/IGF1 signaling pathway [48]. Other researchers found that IGF1 activates the AKT/PKB pathway and protein synthesis pathway by binding to the IGF1 receptor, thereby increasing muscle synthesis [49]. In addition, previous reports have pointed out that mitochondrial dysfunction is the common cause of regulating muscle degeneration induced by aging and T1DM [50]. However, increased oxidative stress, decreased respiration and oxidative capacity, and increased permeability pore opening in mitochondria could promote the process of apoptosis.

Rheumatoid Arthritis and Sarcopenia

RA is an autoimmune disease characterized by the presence of symmetrical polyarthritis, which tends to lead to severe cartilage and bone destruction, thereby resulting in joint pain and arthritis [51]. Persistent joint inflammation can lead to cartilage and bone damage, joint damage, and even disability if the patients

XSL•FO
RenderX

receive treatment not in time [52]. The pathogenesis of RA is complex and has not been fully elucidated [53]. Genetic susceptibility contributes to the increased prevalence of RA pathogenesis [54]. Epigenetic factors, such as DNA methylation and histone acetylation play a key role in RA [55]. The production of autoantibodies and the presence of autoreactive T cells in blood and synovial structures are distinguishing features of RA [56].

Studies have shown that RA is a potentially pathogenic effect of low appendicular lean mass and low grip strength [49]. The injury of RA to skeletal muscle can accelerate the occurrence of sarcopenia and seriously affect the life quality of patients [57,58]. Li et al [59] pointed out that the prevalence of sarcopenia in patients with RA was 31%, which is much higher than that in people without RA. Another cross-sectional study of 388 patients with RA showed that 37.1% of the participants had sarcopenia (14.7% severe sarcopenia, 22.4% sarcopenia) and 49.0% had low muscle mass [60]. Even more importantly, the incidence of sarcopenia increases with the lastingness of RA progression [59,61]. It is worth noting that RA patients with a higher percentage of body fat are more likely to develop sarcopenia [62].

Chronic inflammation is closely related to muscle damage in RA [63]. RA primarily affects the joints, but the risk and susceptibility factors can also cause inflammation in mucosal sites such as the mouth, lungs, and intestines. These sites can be recognized by the adaptive immune system, which triggers an immune response with the help of antigen-presenting cells. This response occurs in secondary lymphoid tissues, resulting in the production of autoantibodies. Subsequently, activated stromal cells like fibroblast-like synovial cells, antigen-presenting cells, and macrophages (MΦ) produce various inflammatory cytokines, including interleukin (IL)-1, IL-2, IL-6, IL-8, IL-17, tumor necrosis factor-α (TNF-α), IFNγ (interferon gamma), and others, which can cause local inflammation and damage [64]. Researchers indicated that skeletal muscle metabolic disorders might be the leading cause of RA combined with sarcopenia [10]. However, skeletal muscle metabolism disorders can lead to loss of skeletal muscle due to impaired muscle anabolism, excessive muscle catabolism, or a combination of both [65]. The ubiquitin-proteasome pathway is a key pathway to increase muscle proteolysis [62,63]. Excessive production of proinflammatory factors makes a significant contribution to the breakdown of muscle protein. Higher levels of TNF-α and IL-1β induced by RA condition could elevate the expression level of IGFBP-5, IGFBP-3, scroggin-1, and muscle ring finger protein-1 (MuRF-1) through regulating the ubiquitin-proteasome system, thereby leading to increased hydrolysis of myofibrillar and soluble proteins [65].

Additionally, it was found that the amount of body fat and the density of muscle were closely linked to disability and physical performance in patients with RA. Specifically, individuals with a higher percentage of body fat showed a stronger correlation with the development of sarcopenia in RA. Therefore, it is recommended that the focus be on reducing fat and improving muscle quality as effective approaches to alleviating the disability of patients with impaired physical functioning [62]. Exercise could increase muscle mass in patients with RA effectively, so physical inactivity might be another risk factor for the high prevalence of RA combined with sarcopenia [66].

Inflammatory Bowel Disease and Sarcopenia

IBD, which includes ulcerative colitis and Crohn disease, is a progressive immune-mediated bowel disease for which there is no effective therapy treatment currently [57,67]. The intestinal microecology of patients with IBD is disturbed, which triggers changes in host immune response and metabolism that subsequently promote inflammation. The inflammatory response to IBD is characterized by a sustained increase in the production of inflammatory cytokines by T cells (TNF-α, IL-6, IL-1b, etc) [68]. Increased IL-12/23 and IFN-c levels, mediated by T helper cell 1, are supposed to be major factors in the maintenance of inflammation in patients with Crohn disease [69]. But T helper cell 2-mediated IL-5 and IL-13 levels elevation is an important factor for inflammation in ulcerative colitis patients [70].

Chronic inflammation and malnutrition resulting from IBD contribute to the development of sarcopenia [71,72]. Sarcopenia is a common condition in IBD patients and it increases fatigue and decreased quality of life. According to a recent study, sarcopenia occurs in 52% of people with Crohn disease and 37% of people with ulcerative colitis [71]. Although the underlying mechanism of sarcopenia in IBD patients remains unclear, studies indicated that the gut-skeletal muscle axis sheds light on the pathophysiology of IBD-associated sarcopenia. This axis comprises specific mediators from the gut that activate signaling pathways for amyotrophic and antimuscular atrophy in skeletal muscle cells [73]. The up-regulation of TNF-α, NF-κB (nuclear factor kappa-B), and JAK/STAT signaling pathways is an important factor in the pathogenesis of sarcopenia [74]. Studies have shown increased levels of proinflammatory factors such as TNF-α, IL-6, circulating lipopolysaccharide, antimyogenic mediators, and promyogenic mediators (myostatin and insulin-like growth factor-1) in IBD patients [75,76]. Both TNF-α and IL-6 can activate NF-κB, which induces transcription of components of the ubiquitin-proteasome proteolytic pathway, such as atrophic protein-1 and MuRF-1, triggering protein degradation [77]. In addition, NF-κB further enhances IL-6 expression, creating a vicious cycle of inflammation and muscle protein degradation [78]. Inhibiting myostatin and muscular dystrophy factors such as atrophin-1 and MuRF-1 and enhancing myogenic factors including MyoG (myogenin) and MyoD (myoblast determination) can improve muscle atrophy [79]. Notably, the JAK/STAT signaling pathway is highly involved in the pathogenesis of IBD [80], mediating the function of several inflammatory cytokines involved in intestinal inflammation, such as IL-2, IL-4, IL-6, IL-7, IL-9, IL-12, IL-15, IL-21, IL-23, and IFN-γ [81].

Malnutrition often occurs in patients with intestinal insufficiency and intestinal failure. Malabsorption of micronutrients and macronutrients resulting from malnutrition can cause changes in body composition and increase the risk of sarcopenia [82]. In general, dietary patterns would have a great influence on the gut microbiota structure and function of patients with inflammatory diseases. For instance, the Mediterranean diet is

XSL•FO
RenderX

beneficial for IBD patients with reduced inflammation and improved immune function, but the Western diet has the opposite effect [83]. In addition, vitamin D is involved in regulating muscle cell proliferation, differentiation, and regeneration. Consistent with this, research has shown that vitamin D supplementation is expected to reduce the occurrence of sarcopenia in IBD patients [84]. The inability to maintain protein balance after meals can lead to a decrease in muscle mass in patients with Crohn disease, exacerbating the occurrence of sarcopenia [85].

Ethical Considerations

This study used exclusively publicly accessible deidentified datasets, ensuring the absence of personal identifiers or sensitive information. Therefore, no ethical approval or declaration was required per the journal's policy and institutional guidelines. Given the absence of secondary analysis involving pre-existing nonidentifiable data, the research maintained full alignment with the Declaration of Helsinki principles. Methodological rigor was further reinforced through strict adherence to ethical protocols outlined in the JMIR Editorial Policies for data governance, as well as compliance with internationally validated open-science frameworks for responsible data use.

Discussion

Sarcopenia is common in the 3 autoimmune diseases commented on in this manuscript, T1DM, RA, and IBD. The combination of autoimmune diseases and sarcopenia can result in a diminished life quality, long-term disease prognosis management, and impose a substantial financial burden. Thus, it is important to investigate the relationship between sarcopenia and autoimmune diseases. Essentially, perturbations within the immune microenvironment of patients with autoimmune disorders elicit fluctuations in the populations of immune cells and alterations in the expression patterns of relevant factors. These modifications subsequently disrupt the functional coherence of muscle regeneration signaling cascades, ultimately underlying the pathogenesis of sarcopenia (Figure 3).

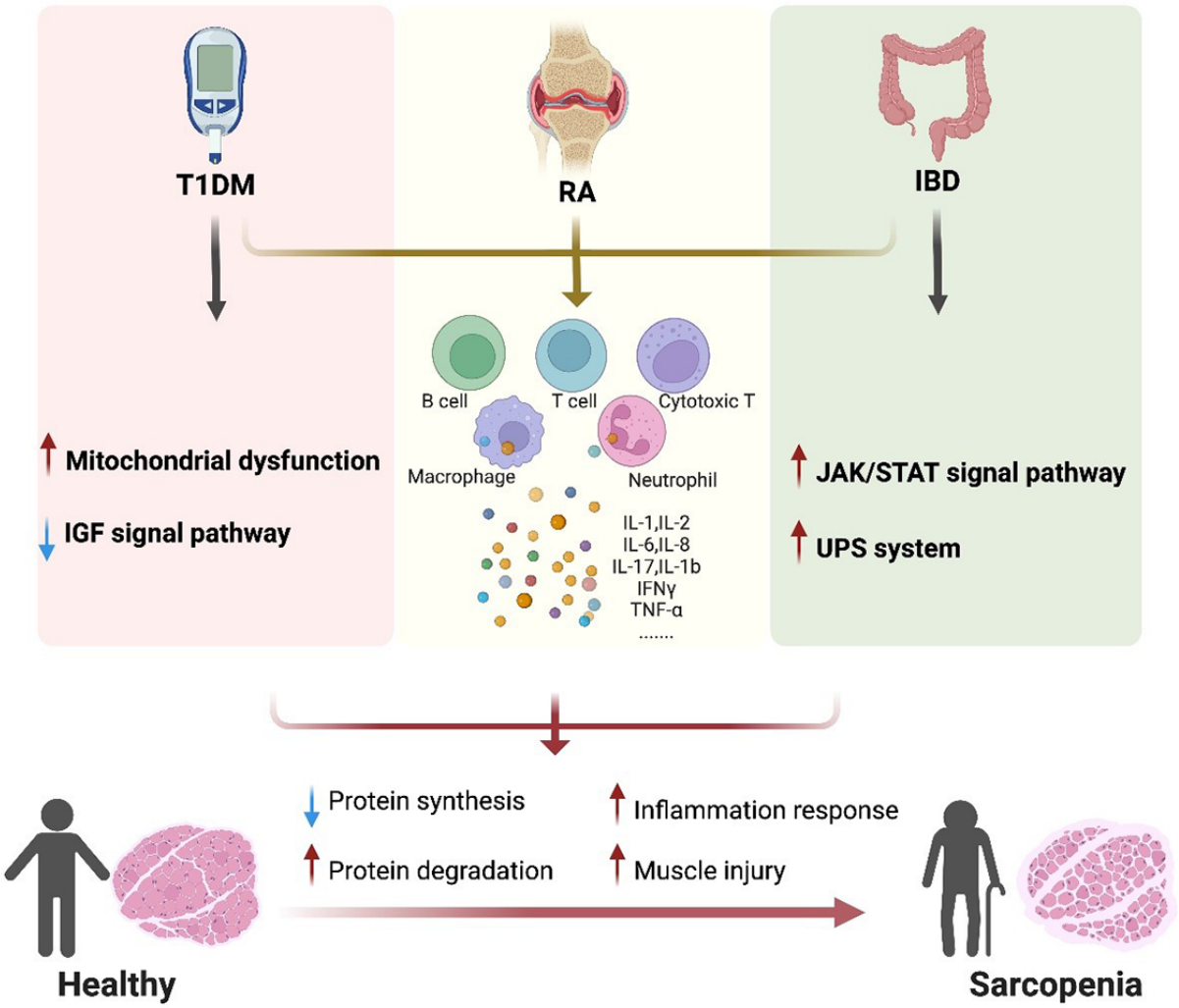
Autoimmune diseases are closely associated with alterations in the immune microenvironment. Pathological interactions between T-cells and B-cells serve as the foundation for numerous autoimmune diseases [86]. For T1DM [87] and RA [88], disruption of T-cell and B-cell tolerance, along with the production of autoantibodies, typically precede the diagnosis of autoimmune diseases. T1DM is characterized as the pancreatic beta cells destroyed by the immune system, resulting in life-long use of exogenous insulin, which places a great burden on patients and medical resources [89]. Pancreatic β cell injury is primarily attributed to T cells while antigen-presenting B cells exacerbate this damage by presenting antigens to T cells [90]. Multiple pieces of evidence indicate that T cells play a pivotal role in driving autoimmune responses in RA, with an abundance of these cells observed within inflamed synovial membranes [91]. The intestinal environment is highly complex, and the immune system maintains the entire intestine's health [92]. Among various factors contributing to the balance of intestinal microbiota, B cells undergo activation and maturation to generate the largest population of plasma cells secreting immunoglobulin A within the human body, thereby promoting immune system homeostasis [93]. Additionally, Follicular helper T cells are enriched in several inflammatory conditions, particularly within the intestine where they exhibit higher levels compared with other tissues [94].

Muscle atrophy represents a significant contributor to sarcopenia development wherein immune cells play a vital role in muscle regeneration. Immune cells release key inflammatory mediators that regulate the muscle fiber microenvironment [95]. Early recruitment of neutrophils leads to robust nonspecific responses that further exacerbate muscle fiber and extracellular matrix injuries [96-98]. Neutrophils exacerbate cellular injury by releasing reactive oxygen species, which induce muscle fiber damage and increase extravasated vascular permeability [99]. Furthermore, neutrophils secrete proinflammatory cytokines IL-6, IL-1β, and TNF-α, promoting satellite cell proliferation while inhibiting differentiation [100]. Both IL-6 and reactive oxygen species have been demonstrated to cause mitochondrial dysfunction/damage and alterations in protein stability within muscle fibers [99,100].

Cytotoxic T lymphocytes exhibit similar functions to neutrophils during the early stages of injury response [101]. Additionally, neutrophils and cytotoxic T lymphocytes recruit inflammatory macrophages to aid in the phagocytosis of damaged tissue for clearance while releasing proinflammatory mediators that stimulate satellite cell and fibroblast proliferation and enhance vascular permeability [102]. However, the investigation of their roles in patients with autoimmune diseases complicated by sarcopenia remains limited. It can be speculated that these disruptions within the skeletal muscle microenvironment may lead to increased loss of muscle mass and fibrosis, resulting in reduced muscle plasticity, weakness, and fatigue.

XSL•FO
RenderX

Figure 3. The potential mechanism of autoimmune diseases combined with sarcopenia. IBD: inflammatory bowel disease; IFN: interferon; IGF: insulin-like growth factor; IL: interleukin; JAK/STAT: Janus kinase/signal transducer and activator of transcription; RA: rheumatoid arthritis; T1DM: type 1 diabetes mellitus; TNF: tumor necrosis factor; UPS: ubiquitin proteasome system.



Conclusions

Autoimmune diseases are closely associated with alterations in the immune microenvironment which plays a crucial role in skeletal muscle differentiation and remodeling [103]. The pathogenesis of autoimmune diseases combined with sarcopenia can be attributed to an imbalance of hormones, growth factors, and proinflammatory factors in the immune microenvironment. This leads to muscle metabolism disorders resulting in a decline in muscle mass due to blocked muscle synthesis and excessive muscle catabolism. To mitigate muscle loss, it is recommended to improve diet quality by consuming nutrients like vitamin D [104] and high-quality protein supplements [105]. Additionally, engaging in physical exercise may be an effective therapy to prevent autoimmune diseases combined with sarcopenia. However, there is still a lack of effective clinical treatment for these diseases, and the pathogenesis of autoimmune diseases combined with sarcopenia needs to be further elucidated [66,106,107]. Future research should focus on understanding the molecular mechanisms of various autoimmune diseases combined with sarcopenia and developing targeted therapies [108].

Acknowledgments

We thank the Gao lab members for the helpful discussion for this study. This study was supported by grants from the Medical and Health Research Project of Zhejiang Province (2025KY1038 and 2024KY182), the Hangzhou Agricultural and Social Development Research Program (20231203A07), the Construction Fund of Medical Key Disciplines of Hangzhou (OO20200055), the Zhejiang Province Science and Technology Plan Project (2022C03035), and the Project of Hangzhou Health Science and Technology Program (A20200210).

XSL•FO
RenderX

Authors' Contributions

YZ, SH, and YG designed the study. YZ, Q Wu, and Yi Wang wrote the manuscript. QC, BL, QZ, Q Wang, Yule Wang, SH, and YG provided comments and revised the manuscript. Yi Wang, SH, and YG supervised the study. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Chen LK, Woo J, Assantachai P, et al. Asian working group for sarcopenia: 2019 consensus update on sarcopenia diagnosis and treatment. J Am Med Dir Assoc 2020 Mar;21(3):300-307. [doi: 10.1016/j.jamda.2019.12.012] [Medline: 32033882]
2. Sayer AA, Cruz-Jentoft A. Sarcopenia definition, diagnosis and treatment: consensus is growing. Age Ageing 2022 Oct 6;51(10):10. [doi: 10.1093/ageing/afac220]
3. Lipina C, Hundal HS. Lipid modulation of skeletal muscle mass and function. J Cachexia Sarcopenia Muscle 2017 Apr;8(2):190-201. [doi: 10.1002/jcsm.12144]
4. Kitamura A, Seino S, Abe T, et al. Sarcopenia: prevalence, associated factors, and the risk of mortality and disability in Japanese older adults. J Cachexia Sarcopenia Muscle 2021 Feb;12(1):30-38. [doi: 10.1002/jcsm.12651] [Medline: 33241660]
5. Ou Q, Pan D. Sarcopenia and risk of cardio-cerebrovascular disease: a two-sample Mendelian randomization study. BST 2023;17(5):413-414. [doi: 10.5582/bst.2023.01246]
6. Eitmann S, Matrai P, Hegyi P, et al. Obesity paradox in older sarcopenic adults - a delay in aging: a systematic review and meta-analysis. Ageing Res Rev 2024 Jan;93:102164. [doi: 10.1016/j.arr.2023.102164] [Medline: 38103840]
7. Jones G, Trajanoska K, Santanasto AJ, et al. Genome-wide meta-analysis of muscle weakness identifies 15 susceptibility loci in older men and women. Nat Commun ;12(1):654. [doi: 10.1038/s41467-021-20918-w]
8. Liu S, D'Amico D, Shankland E, et al. Effect of urolithin a supplementation on muscle endurance and mitochondrial health in older adults: a randomized clinical trial. JAMA Netw Open 2022 Jan 4;5(1):e2144279. [doi: 10.1001/jamanetworkopen.2021.44279] [Medline: 35050355]
9. Tsuzuku S, Kajioka T, Sakakibara H, Shimaoka K. Slow movement resistance training using body weight improves muscle mass in the elderly: a randomized controlled trial. Scand J Med Sci Sports 2018 Apr;28(4):1339-1344. [doi: 10.1111/sms.13039] [Medline: 29247985]
10. Riuzzi F, Sorci G, Arcuri C, et al. Cellular and molecular mechanisms of sarcopenia: the S100B perspective. J Cachexia Sarcopenia Muscle 2018 Dec;9(7):1255-1268. [doi: 10.1002/jcsm.12363]
11. An HJ, Tizaoui K, Terrazzino S, et al. Sarcopenia in autoimmune and rheumatic diseases: a comprehensive review. Int J Mol Sci 2020 Aug 7;21(16):16. [doi: 10.3390/ijms21165678] [Medline: 32784808]
12. Balestrieri P, Ribolsi M, Guarino MPL, Emerenziani S, Altomare A, Cicala M. Nutritional aspects in inflammatory bowel diseases. Nutrients 2020 Jan 31;12(2):372. [doi: 10.3390/nu12020372] [Medline: 32023881]
13. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing 2019 Jan 1;48(1):16-31. [doi: 10.1093/ageing/afy169] [Medline: 30312372]
14. Vikberg S, Sörlén N, Brandén L, et al. Effects of resistance training on functional strength and muscle mass in 70-year-old individuals with pre-sarcopenia: a randomized controlled trial. J Am Med Dir Assoc 2019 Jan;20(1):28-34. [doi: 10.1016/j.jamda.2018.09.011]
15. Hua H, Xu X, Tang Y, Ren Z, Xu Q, Chen L. Effect of sarcopenia on clinical outcomes following digestive carcinoma surgery: a meta-analysis. Support Care Cancer 2019 Jul;27(7):2385-2394. [doi: 10.1007/s00520-019-04767-4]
16. Cho MR, Lee S, Song SK. A review of sarcopenia pathophysiology, diagnosis, treatment and future direction. J Korean Med Sci 2022 May 9;37(18):e146. [doi: 10.3346/jkms.2022.37.e146] [Medline: 35535373]
17. Beaudart C, McCloskey E, Bruyère O, et al. Sarcopenia in daily practice: assessment and management. BMC Geriatr 2016 Oct 5;16(1):170. [doi: 10.1186/s12877-016-0349-4] [Medline: 27716195]
18. Messina C, Maffi G, Vitale JA, Ulivieri FM, Guglielmi G, Sconfienza LM. Diagnostic imaging of osteoporosis and sarcopenia: a narrative review. Quant Imaging Med Surg 2018 Feb;8(1):86-99. [doi: 10.21037/qims.2018.01.01]
19. Mueller L, Mentil N, Staub N, et al. Association of thoracic skeletal muscle index with clinical outcome and response to nutritional interventions in patients at risk of malnutrition-secondary analysis of a randomized trial. Nutrients 2023 Feb 5;15(4):817. [doi: 10.3390/nu15040817] [Medline: 36839175]
20. Mijnarends DM, Meijers JMM, Halfens RJG, et al. Validity and reliability of tools to measure muscle mass, strength, and physical performance in community-dwelling older people: a systematic review. J Am Med Dir Assoc 2013 Mar;14(3):170-178. [doi: 10.1016/j.jamda.2012.10.009] [Medline: 23276432]
21. Albano D, Messina C, Vitale J, Sconfienza LM. Imaging of sarcopenia: old evidence and new insights. Eur Radiol 2020 Apr;30(4):2199-2208. [doi: 10.1007/s00330-019-06573-2]
22. Kyle UG, Bosaeus I, De Lorenzo AD, et al. Bioelectrical impedance analysis--part I: review of principles and methods. Clin Nutr 2004 Oct;23(5):1226-1243. [doi: 10.1016/j.clnu.2004.06.004] [Medline: 15380917]

XSL•FO
RenderX

23. Gonzalez MC, Barbosa-Silva TG, Heymsfield SB. Bioelectrical impedance analysis in the assessment of sarcopenia. Curr Opin Clin Nutr Metab Care 2018 Sep;21(5):366-374. [doi: 10.1097/MCO.0000000000000496] [Medline: 29957677]
24. Bosaeus I, Wilcox G, Rothenberg E, Strauss BJ. Skeletal muscle mass in hospitalized elderly patients: comparison of measurements by single-frequency BIA and DXA. Clin Nutr 2014 Jun;33(3):426-431. [doi: 10.1016/j.clnu.2013.06.007]
25. Reiss J, Iglseder B, Kreutzer M, et al. Case finding for sarcopenia in geriatric inpatients: performance of bioimpedance analysis in comparison to dual x-ray absorptiometry. BMC Geriatr 2016 Feb 29;16:52. [doi: 10.1186/s12877-016-0228-z] [Medline: 26928275]
26. Hacker UT, Hasenclever D, Baber R, et al. Modified Glasgow prognostic score (mGPS) is correlated with sarcopenia and dominates the prognostic role of baseline body composition parameters in advanced gastric and esophagogastric junction cancer patients undergoing first-line treatment from the phase III EXPAND trial. Ann Oncol 2022 Jul;33(7):685-692. [doi: 10.1016/j.annonc.2022.03.274] [Medline: 35395383]
27. Shivaji UN, Nardone OM, Cannatelli R, Smith SC, Ghosh S, Iacucci M. Small molecule oral targeted therapies in ulcerative colitis. Lancet Gastroenterol Hepatol 2020 Sep;5(9):850-861. [doi: 10.1016/S2468-1253(19)30414-5] [Medline: 32171056]
28. Wang CR, Tsai HW. Autoimmune liver diseases in systemic rheumatic diseases. World J Gastroenterol 2022 Jun 21;28(23):2527-2545. [doi: 10.3748/wjg.v28.i23.2527] [Medline: 35949355]
29. Cao F, Liu YC, Ni QY, et al. Temporal trends in the prevalence of autoimmune diseases from 1990 to 2019. Autoimmun Rev 2023 Aug;22(8):103359. [doi: 10.1016/j.autrev.2023.103359] [Medline: 37201621]
30. Wang L, Wang FS, Gershwin ME. Human autoimmune diseases: a comprehensive update. J Intern Med 2015 Oct;278(4):369-395. [doi: 10.1111/joim.12395] [Medline: 26212387]
31. Miller FW. The increasing prevalence of autoimmunity and autoimmune diseases: an urgent call to action for improved understanding, diagnosis, treatment, and prevention. Curr Opin Immunol 2023 Feb;80:102266. [doi: 10.1016/j.coi.2022.102266]
32. Akil AAS, Yassin E, Al-Maraghi A, Aliyev E, Al-Malki K, Fakhro KA. Diagnosis and treatment of type 1 diabetes at the dawn of the personalized medicine era. J Transl Med 2021 Apr 1;19(1):137. [doi: 10.1186/s12967-021-02778-6] [Medline: 33794915]
33. Wu D, Luo Y, Li T, et al. Systemic complications of rheumatoid arthritis: focus on pathogenesis and treatment. Front Immunol 2022;13:1051082. [doi: 10.3389/fimmu.2022.1051082]
34. Singh N, Bernstein CN. Environmental risk factors for inflammatory bowel disease. UEG Journal 2022 Dec;10(10):1047-1053. [doi: 10.1002/ueg2.12319]
35. Massironi S, Viganò C, Palermo A, et al. Inflammation and malnutrition in inflammatory bowel disease. Lancet Gastroenterol Hepatol 2023 Jun;8(6):579-590. [doi: 10.1016/S2468-1253(23)00011-0]
36. Argollo M, Gilardi D, Peyrin-Biroulet C, Chabot JF, Peyrin-Biroulet L, Danese S. Comorbidities in inflammatory bowel disease: a call for action. Lancet Gastroenterol Hepatol 2019 Aug;4(8):643-654. [doi: 10.1016/S2468-1253(19)30173-6] [Medline: 31171484]
37. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. Nat Rev Nephrol 2020 Jul;16(7):377-390. [doi: 10.1038/s41581-020-0278-5] [Medline: 32398868]
38. Figus FA, Piga M, Azzolin I, McConnell R, Iagnocco A. Rheumatoid arthritis: extra-articular manifestations and comorbidities. Autoimmun Rev 2021 Apr;20(4):102776. [doi: 10.1016/j.autrev.2021.102776] [Medline: 33609792]
39. Herold KC, Delong T, Perdigoto AL, Biru N, Brusko TM, Walker LSK. The immunology of type 1 diabetes. Nat Rev Immunol 2024 Jun;24(6):435-451. [doi: 10.1038/s41577-023-00985-4]
40. Eizirik DL, Pasquali L, Cnop M. Pancreatic β-cells in type 1 and type 2 diabetes mellitus: different pathways to failure. Nat Rev Endocrinol 2020 Jul;16(7):349-362. [doi: 10.1038/s41574-020-0355-7] [Medline: 32398822]
41. Sadeghi M, Dehnavi S, Jamialahmadi T, Johnston TP, Sahebkar A. Neutrophil extracellular trap: a key player in the pathogenesis of autoimmune diseases. Int Immunopharmacol 2023 Mar;116:109843. [doi: 10.1016/j.intimp.2023.109843] [Medline: 36764274]
42. Maratova K, Soucek O, Matyskova J, et al. Muscle functions and bone strength are impaired in adolescents with type 1 diabetes. Bone 2018 Jan;106:22-27. [doi: 10.1016/j.bone.2017.10.005]
43. Mori H, Kuroda A, Yoshida S, et al. High prevalence and clinical impact of dynapenia and sarcopenia in Japanese patients with type 1 and type 2 diabetes: findings from the impact of diabetes mellitus on dynapenia study. J of Diabetes Invest 2021 Jun;12(6):1050-1059. [doi: 10.1111/jdi.13436]
44. Sylow L, Tokarz VL, Richter EA, Klip A. The many actions of insulin in skeletal muscle, the paramount tissue determining glycemia. Cell Metab 2021 Apr 6;33(4):758-780. [doi: 10.1016/j.cmet.2021.03.020] [Medline: 33826918]
45. Hiromine Y, Noso S, Rakugi H, et al. Poor glycemic control rather than types of diabetes is a risk factor for sarcopenia in diabetes mellitus: the MUSCLES - DM study. J of Diabetes Invest 2022 Nov;13(11):1881-1888. [doi: 10.1111/jdi.13882]
46. Shapiro MR, Wasserfall CH, McGrail SM, et al. Insulin-like growth factor dysregulation both preceding and following type 1 diabetes diagnosis. Diabetes 2020 Mar 1;69(3):413-423. [doi: 10.2337/db19-0942]
47. Hata S, Mori H, Yasuda T, et al. A low serum IGF-1 is correlated with sarcopenia in subjects with type 1 diabetes mellitus: findings from a post-hoc analysis of the iDIAMOND study. Diabetes Res Clin Pract 2021 Sep;179:108998. [doi: 10.1016/j.diabres.2021.108998] [Medline: 34390761]

XSL•FO
RenderX

48. Perrini S, Laviola L, Carreira MC, Cignarelli A, Natalicchio A, Giorgino F. The GH/IGF1 axis and signaling pathways in the muscle and bone: mechanisms underlying age-related skeletal muscle wasting and osteoporosis. J Endocrinol 2010 Jun;205(3):201-210. [doi: 10.1677/JOE-09-0431] [Medline: 20197302]
49. Schiaffino S, Mammucari C. Regulation of skeletal muscle growth by the IGF1-Akt/PKB pathway: insights from genetic models. Skelet Muscle 2011 Jan 24;1(1):4. [doi: 10.1186/2044-5040-1-4] [Medline: 21798082]
50. Alway SE. Mitochondrial dysfunction: linking type 1 diabetes and sarcopenia. Exerc Sport Sci Rev 2019 Apr;47(2):63. [doi: 10.1249/JES.0000000000000186] [Medline: 30882453]
51. Krainer J, Siebenhandl S, Weinhäusel A. Systemic autoinflammatory diseases. J Autoimmun 2020 May;109:102421. [doi: 10.1016/j.jaut.2020.102421] [Medline: 32019685]
52. Pietschmann P, Butylina M, Kerschan-Schindl K, Sipos W. Mechanisms of systemic osteoporosis in rheumatoid arthritis. Int J Mol Sci 2022 Aug 5;23(15):15. [doi: 10.3390/ijms23158740] [Medline: 35955873]
53. Gravallese EM, Firestein GS. Rheumatoid arthritis - common origins, divergent mechanisms. N Engl J Med 2023 Feb 9;388(6):529-542. [doi: 10.1056/NEJMra2103726] [Medline: 36780677]
54. Hemminki K, Li X, Sundquist J, Sundquist K. Familial associations of rheumatoid arthritis with autoimmune diseases and related conditions. Arthritis Rheum 2009 Mar;60(3):661-668. [doi: 10.1002/art.24328] [Medline: 19248111]
55. Nemtsova MV, Zaletaev DV, Bure IV, et al. Epigenetic changes in the pathogenesis of rheumatoid arthritis. Front Genet 2019;10:570. [doi: 10.3389/fgene.2019.00570] [Medline: 31258550]
56. Di Matteo A, Bathon JM, Emery P. Rheumatoid arthritis. Lancet 2023 Nov 25;402(10416):2019-2033. [doi: 10.1016/S0140-6736(23)01525-8] [Medline: 38240831]
57. Bennett JL, Pratt AG, Dodds R, Sayer AA, Isaacs JD. Rheumatoid sarcopenia: loss of skeletal muscle strength and mass in rheumatoid arthritis. Nat Rev Rheumatol 2023 Apr;19(4):239-251. [doi: 10.1038/s41584-023-00921-9] [Medline: 36801919]
58. Fukuda W, Omoto A, Ohta T, et al. Low body mass index is associated with impaired quality of life in patients with rheumatoid arthritis. Int J Rheum Dis 2013 Jun;16(3):297-302. [doi: 10.1111/1756-185X.12079] [Medline: 23981751]
59. Li TH, Chang YS, Liu CW, et al. The prevalence and risk factors of sarcopenia in rheumatoid arthritis patients: a systematic review and meta-regression analysis. Semin Arthritis Rheum 2021 Feb;51(1):236-245. [doi: 10.1016/j.semarthrit.2020.10.002] [Medline: 33385864]
60. Torii M, Hashimoto M, Hanai A, et al. Prevalence and factors associated with sarcopenia in patients with rheumatoid arthritis. Mod Rheumatol 2019 Jul;29(4):589-595. [doi: 10.1080/14397595.2018.1510565] [Medline: 30092163]
61. Dietzel R, Wiegmann S, Borucki D, et al. Prevalence of sarcopenia in patients with rheumatoid arthritis using the revised EWGSOP2 and the FNIH definition. RMD Open 2022 Sep;8(2):2. [doi: 10.1136/rmdopen-2022-002600] [Medline: 36180102]
62. Vlietstra L, Stebbings S, Meredith-Jones K, Abbott JH, Treharne GJ, Waters DL. Sarcopenia in osteoarthritis and rheumatoid arthritis: the association with self-reported fatigue, physical function and obesity. PLoS ONE 2019;14(6):e0217462. [doi: 10.1371/journal.pone.0217462] [Medline: 31170172]
63. Andonian BJ, Huffman KM. Skeletal muscle disease in rheumatoid arthritis: the center of cardiometabolic comorbidities? Curr Opin Rheumatol 2020 May;32(3):297-306. [doi: 10.1097/BOR.0000000000000697] [Medline: 32141950]
64. Smolen JS, Aletaha D, Barton A, et al. Rheumatoid arthritis. Nat Rev Dis Primers 2018 Feb 8;4:18001. [doi: 10.1038/nrdp.2018.1] [Medline: 29417936]
65. Castillero E, Martín AI, López-Menduiña M, Granado M, Villanúa MA, López-Calderón A. IGF-I system, atrogenes and myogenic regulatory factors in arthritis induced muscle wasting. Mol Cell Endocrinol 2009 Oct 15;309(1-2):8-16. [doi: 10.1016/j.mce.2009.05.017] [Medline: 19501629]
66. Liao CD, Chen HC, Huang SW, Liou TH. Exercise therapy for sarcopenia in rheumatoid arthritis: a meta-analysis and meta-regression of randomized controlled trials. Clin Rehabil 2022 Feb;36(2):145-157. [doi: 10.1177/02692155211035539] [Medline: 34404254]
67. Chazaud B. Inflammation and skeletal muscle regeneration: leave it to the macrophages!. Trends Immunol 2020 Jun;41(6):481-492. [doi: 10.1016/j.it.2020.04.006] [Medline: 32362490]
68. Boirivant M, Cossu A. Inflammatory bowel disease. Oral Dis 2012 Jan;18(1):1-15. [doi: 10.1111/j.1601-0825.2011.01811.x] [Medline: 21564424]
69. Fuss IJ, Heller F, Boirivant M, et al. Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. J Clin Invest 2004 May;113(10):1490-1497. [doi: 10.1172/JCI19836] [Medline: 15146247]
70. Bouma G, Strober W. The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol 2003 Jul;3(7):521-533. [doi: 10.1038/nri1132] [Medline: 12876555]
71. Dhaliwal A, Quinlan JI, Overthrow K, et al. Sarcopenia in inflammatory bowel disease: a narrative overview. Nutrients 2021 Feb 17;13(2):656. [doi: 10.3390/nu13020656] [Medline: 33671473]
72. Steell L, Gray SR, Russell RK, et al. Pathogenesis of musculoskeletal deficits in children and adults with inflammatory bowel disease. Nutrients 2021 Aug 23;13(8):2899. [doi: 10.3390/nu13082899] [Medline: 34445056]
73. Ehlers L, Bannert K, Rohde S, et al. Preclinical insights into the gut-skeletal muscle axis in chronic gastrointestinal diseases. J Cell Mol Med 2020 Aug;24(15):8304-8314. [doi: 10.1111/jcmm.15554] [Medline: 32628812]

XSL•FO
RenderX

74. Nardone OM, de Sire R, Petito V, et al. Inflammatory bowel diseases and sarcopenia: the role of inflammation and gut microbiota in the development of muscle failure. Front Immunol 2021;12:694217. [doi: 10.3389/fimmu.2021.694217] [Medline: 34326845]
75. de Sire R, Rizzatti G, Ingravalle F, et al. Skeletal muscle-gut axis: emerging mechanisms of sarcopenia for intestinal and extra intestinal diseases. Minerva Gastroenterol Dietol 2018 Dec;64(4):351-362. [doi: 10.23736/S1121-421X.18.02511-4]
76. Bian AL, Hu HY, Rong YD, Wang J, Wang JX, Zhou XZ. A study on relationship between elderly sarcopenia and inflammatory factors IL-6 and TNF-α. Eur J Med Res 2017 Jul 12;22(1):25. [doi: 10.1186/s40001-017-0266-9] [Medline: 28701179]
77. Lin SY, Chen WY, Lee FY, Huang CJ, Sheu WHH. Activation of ubiquitin-proteasome pathway is involved in skeletal muscle wasting in a rat model with biliary cirrhosis: potential role of TNF-alpha. Am J Physiol Endocrinol Metab 2005 Mar;288(3):E493-E501. [doi: 10.1152/ajpendo.00186.2004] [Medline: 15522995]
78. Vella L, Caldow MK, Larsen AE, et al. Resistance exercise increases NF-κB activity in human skeletal muscle. Am J Physiol Regul Integr Comp Physiol 2012 Mar 15;302(6):R667-R673. [doi: 10.1152/ajpregu.00336.2011] [Medline: 22189669]
79. Hong Y, Lee JH, Jeong KW, Choi CS, Jun HS. Amelioration of muscle wasting by glucagon-like peptide-1 receptor agonist in muscle atrophy. J Cachexia Sarcopenia Muscle 2019 Aug;10(4):903-918. [doi: 10.1002/jcsm.12434] [Medline: 31020810]
80. Olivera PA, Lasa JS, Bonovas S, Danese S, Peyrin-Biroulet L. Safety of janus kinase inhibitors in patients with inflammatory bowel diseases or other immune-mediated diseases: a systematic review and meta-analysis. Gastroenterology 2020 May;158(6):1554-1573. [doi: 10.1053/j.gastro.2020.01.001] [Medline: 31926171]
81. Liu C, Liu N, Xia Y, Zhao Z, Xiao T, Li H. Osteoporosis and sarcopenia-related traits: a bi-directional Mendelian randomization study. Front Endocrinol (Lausanne) 2022;13:975647. [doi: 10.3389/fendo.2022.975647] [Medline: 36187130]
82. Skallerup A, Nygaard L, Olesen SS, Køhler M, Vinter-Jensen L, Rasmussen HH. The prevalence of sarcopenia is markedly increased in patients with intestinal failure and associates with several risk factors. Clin Nutr 2018 Dec;37(6 Pt A):2029-2035. [doi: 10.1016/j.clnu.2017.09.010] [Medline: 29029893]
83. Aziz T, Khan AA, Tzora A, Voidarou CC, Skoufos I. Dietary implications of the bidirectional relationship between the gut microflora and inflammatory diseases with special emphasis on irritable bowel disease: current and future perspective. Nutrients 2023 Jun 29;15(13):2956. [doi: 10.3390/nu15132956] [Medline: 37447285]
84. Palmese F, Del Toro R, Di Marzio G, Cataleta P, Sama MG, Domenicali M. Sarcopenia and vitamin D deficiency in patients with Crohn's disease: pathological conditions that should be linked together. Nutrients 2021 Apr 20;13(4):1378. [doi: 10.3390/nu13041378]
85. Pizzoferrato M, de Sire R, Ingravalle F, et al. Characterization of sarcopenia in an IBD population attending an Italian gastroenterology tertiary center. Nutrients 2019 Sep 24;11(10):2281. [doi: 10.3390/nu11102281] [Medline: 31554166]
86. Marks KE, Rao DA. T peripheral helper cells in autoimmune diseases. Immunol Rev 2022 May;307(1):191-202. [doi: 10.1111/imr.13069] [Medline: 35103314]
87. Sosenko JM, Skyler JS, Herold KC, Palmer JP, Type 1 Diabetes TrialNet and Diabetes Prevention Trial–Type 1 Study Groups. The metabolic progression to type 1 diabetes as indicated by serial oral glucose tolerance testing in the diabetes prevention trial-type 1. Diabetes 2012 Jun;61(6):1331-1337. [doi: 10.2337/db11-1660] [Medline: 22618768]
88. Nielen MMJ, van Schaardenburg D, Reesink HW, et al. Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors. Arthritis & Rheumatism 2004 Feb;50(2):380-386. [doi: 10.1002/art.20018]
89. Jeker LT, Bour-Jordan H, Bluestone JA. Breakdown in peripheral tolerance in type 1 diabetes in mice and humans. Cold Spring Harb Perspect Med 2012 Mar;2(3):a007807. [doi: 10.1101/cshperspect.a007807] [Medline: 22393537]
90. Smith MJ, Simmons KM, Cambier JC. B cells in type 1 diabetes mellitus and diabetic kidney disease. Nat Rev Nephrol 2017 Oct 17;13(11):712-720. [doi: 10.1038/nrneph.2017.138] [Medline: 29038537]
91. Diogo D, Okada Y, Plenge RM. Genome-wide association studies to advance our understanding of critical cell types and pathways in rheumatoid arthritis: recent findings and challenges. Curr Opin Rheumatol 2014 Jan;26(1):85-92. [doi: 10.1097/BOR.0000000000000012] [Medline: 24276088]
92. Wastyk HC, Fragiadakis GK, Perelman D, et al. Gut-microbiota-targeted diets modulate human immune status. Cell 2021 Aug 5;184(16):4137-4153. [doi: 10.1016/j.cell.2021.06.019] [Medline: 34256014]
93. Lundell AC, Björnsson V, Ljung A, et al. Infant B cell memory differentiation and early gut bacterial colonization. J Immunol 2012 May 1;188(9):4315-4322. [doi: 10.4049/jimmunol.1103223] [Medline: 22490441]
94. Spencer J, Bemark M. Human intestinal B cells in inflammatory diseases. Nat Rev Gastroenterol Hepatol 2023 Apr;20(4):254-265. [doi: 10.1038/s41575-023-00755-6] [Medline: 36849542]
95. Tu H, Li YL. Inflammation balance in skeletal muscle damage and repair. Front Immunol 2023;14:1133355. [doi: 10.3389/fimmu.2023.1133355] [Medline: 36776867]
96. Stackowicz J, Jönsson F, Reber LL. Mouse models and tools for the in vivo study of neutrophils. Front Immunol 2019;10:3130. [doi: 10.3389/fimmu.2019.03130]
97. Xiao Y, Cheng Y, Liu WJ, et al. Effects of neutrophil fate on inflammation. Inflamm Res 2023 Dec;72(12):2237-2248. [doi: 10.1007/s00011-023-01811-2] [Medline: 37925664]

XSL•FO
RenderX

98. Singhal A, Kumar S. Neutrophil and remnant clearance in immunity and inflammation. Immunology 2022 Jan;165(1):22-43. [doi: 10.1111/imm.13423] [Medline: 34704249]
99. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. Nat Rev Immunol 2013 Mar;13(3):159-175. [doi: 10.1038/nri3399] [Medline: 23435331]
100. Mayadas TN, Cullere X, Lowell CA. The multifaceted functions of neutrophils. Annu Rev Pathol 2014;9:181-218. [doi: 10.1146/annurev-pathol-020712-164023] [Medline: 24050624]
101. Deyhle MR, Hyldahl RD. The role of T lymphocytes in skeletal muscle repair from traumatic and contraction-induced injury. Front Physiol 2018;9:768. [doi: 10.3389/fphys.2018.00768] [Medline: 29973887]
102. VanderVeen BN, Murphy EA, Carson JA. The impact of immune cells on the skeletal muscle microenvironment during cancer cachexia. Front Physiol 2020;11:1037. [doi: 10.3389/fphys.2020.01037] [Medline: 32982782]
103. Morgan J, Partridge T. Skeletal muscle in health and disease. Dis Model Mech 2020 Feb 6;13(2):dmm042192. [doi: 10.1242/dmm.042192] [Medline: 32066552]
104. Alliband KH, Kozhevnikova SV, Parr T, Jethwa PH, Brameld JM. In vitro effects of biologically active vitamin D on myogenesis: a systematic review. Front Physiol 2021;12:736708. [doi: 10.3389/fphys.2021.736708] [Medline: 34566700]
105. Ahmadi A, Eftekhari MH, Mazloom Z, et al. Fortified whey beverage for improving muscle mass in chronic obstructive pulmonary disease: a single-blind, randomized clinical trial. Respir Res 2020 Aug 17;21(1):216. [doi: 10.1186/s12931-020-01466-1] [Medline: 32807165]
106. Mc Gettigan N, Allen K, Saeidi R, O' Toole A, Boland K. A systematic review of the effect of structured exercise on inflammation and body composition in inflammatory bowel disease. Int J Colorectal Dis 2023 May 25;38(1):143. [doi: 10.1007/s00384-023-04437-2] [Medline: 37227593]
107. Harmer AR, Chisholm DJ, McKenna MJ, et al. Sprint training increases muscle oxidative metabolism during high-intensity exercise in patients with type 1 diabetes. Diabetes Care 2008 Nov;31(11):2097-2102. [doi: 10.2337/dc08-0329] [Medline: 18716051]
108. Pascual-Fernández J, Fernández-Montero A, Córdova-Martínez A, Pastor D, Martínez-Rodríguez A, Roche E. Sarcopenia: molecular pathways and potential targets for intervention. Int J Mol Sci 2020 Nov 22;21(22):22. [doi: 10.3390/ijms21228844] [Medline: 33266508]

Abbreviations

BIA: bioelectrical impedance analysis
DXA: dual-energy x-ray absorptiometry
IBD: inflammatory bowel disease
IFNγ: interferon gamma
IGF: insulin-like growth factor
IGFBP: insulin-like growth factor–binding protein
IL: interleukin
JAK/STAT: Janus kinase/signal transducers and activators of transcription
MuRF-1: muscle ring finger protein-1
MyoD: myoblast determination
MyoG: myogenin
NF-κB: nuclear factor kappa-B
RA: rheumatoid arthritis
T1DM: type 1 diabetes mellitus
TNF-α: tumor necrosis factor-α

Edited by TDA Cardoso; submitted 17.07.24; peer-reviewed by E Antuna, M Gasmi, YY Kristian; revised version received 25.02.25; accepted 25.02.25; published 23.05.25.

Please cite as:
Zhang Y, Wu Q, Wang Y, Chen Q, Han S, Li B, Zhao Q, Wang Q, Wang Y, Gao Y
Systemic Inflammation and Disruption of the Local Microenvironment Compromise Muscle Regeneration: Critical Pathogenesis of Autoimmune-Associated Sarcopenia
Interact J Med Res 2025;14:e64456
URL: https://www.i-jmr.org/2025/1/e64456
doi:10.2196/64456

XSL•FO
RenderX

© Yingjuan Zhang, Qingqian Wu, Yi Wang, Qingyan Chen, Shuang Han, Bei Li, Qingwen Zhao, Qianzhuo Wang, Yule Wang, Yue Gao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 23.5.2025. 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 to Design Electronic Case Report Form (eCRF) Questions to Maximize Semantic Interoperability in Clinical Research

Eugenia Rinaldi*, MSc; Caroline Stellmach*, MSc; Sylvia Thun, MD

Core Unit Digital Medicine and Interoperability, Berlin Institute of Health, Charité - Universitätsmedizin Berlin, Luisenstr. 65, Berlin, Germany
*these authors contributed equally

Corresponding Author:
Eugenia Rinaldi, MSc
Core Unit Digital Medicine and Interoperability, Berlin Institute of Health, Charité - Universitätsmedizin Berlin, Luisenstr. 65, Berlin, Germany

Abstract

Case report forms (CRFs) are the instruments used by research organizations worldwide to collect information about patients and study participants with the purpose of answering specific questions, assessing the efficacy and safety of medical products, and in general improving prevention and treatment in health care. To obtain significant research results out of the collected data, CRFs should be designed following the recommendations issued by regulatory authorities. However, we believe that semantic interoperability in CRFs has not yet been properly addressed. Within an international consortium comprising several COVID-19 cohorts, we scrutinized the questions included in the different CRFs with the purpose of establishing semantic interoperability across the different study data elements so that data could be merged and jointly analyzed. We realized that similar concepts were structured very differently across the different CRFs, making it hard to find and match the information. Based on the experience acquired, we developed 5 guiding principles on how to design CRFs to support semantic interoperability and increase data quality while also facilitating the sharing of data. Our aim in this viewpoint is to provide general suggestions that, in our opinion, should support researchers in designing CRFs. We conclude by urging authorities to establish an international coordination board for standards and interoperable clinical study data with competence in clinical data, interoperability standards, and data protection as part of a preparedness plan for future pandemics or other health threats.

(Interact J Med Res 2025;14:e51598) doi:10.2196/51598

KEYWORDS

case report form; CRF; interoperability; standard data model; data format; metadata; core data elements; data quality

Introduction

Since the onset of the COVID-19 pandemic, we have witnessed the emergence of numerous studies worldwide aimed at deepening our understanding of SARS-CoV-2 infection and enhancing treatment strategies [1]. The urgency to produce meaningful research results in a relatively short timeframe has underscored the importance of efficiently merging data from diverse studies. However, differences in languages, formats, and terminologies often complicate the sharing and integration of data.

The European project ORCHESTRA [2], which sought to build a pan-European cohort of COVID-19 patients, confronted the significant challenge of harmonizing data from various cross-country studies [3]. Achieving interoperability among disparate datasets is critical not only for supporting ongoing research but also for bolstering preparedness against future global health crises. This aligns with the European Commission's objectives under the European Health Data Space regulation [4], which emphasizes improving data accessibility and integration across member states.

Case report forms (CRFs) are essential tools employed by research organizations worldwide to gather detailed information from patients and study participants. These forms are designed to address specific research questions, evaluate the efficacy and safety of medical interventions, and ultimately advance prevention and treatment in health care.

A clinical study is designed to answer one or several research questions based on the analysis of data collected from patients that have been enrolled and are being observed or are partaking in specific interventions following the study protocol. CRFs, either in paper form or electronic format, are the instruments used to collect study data and form the basis of any subsequent statistical analysis. Electronic CRFs (eCRFs) are preferred over paper CRFs because data can automatically be stored in a digital format and immediately used, removing inaccuracies derived from the interpretation and transcription of handwriting. Additionally, if properly structured, the digital format intrinsically offers great potential for data objects to be more findable, accessible, interoperable, and reusable, in other words, more FAIR [5], than paper. The design of eCRFs is crucial for the outcome of a study [6]. Therefore, it should be optimized to enhance data quality and data interoperability.

XSL•FO
RenderX

Ideally, a standard operating procedure is established initially for designing eCRFs. Common recommendations for eCRF design include suggestions to reduce data entry errors and ambiguity in the interpretation of variables such as maximizing the use of coded questions and answer lists and minimizing the use of free text answers; using built-in consistency checks for admissible ranges and plausible date checks; facilitating data entry using branching logic strategies; specifying units of measurement (particularly for laboratory parameters, but also for vital signs, etc); adopting standard data formats; using (and reusing) published Common Data Elements, if available, and unambiguous temporal reference (eg, before or during infection).

Regulatory authorities and international expert organizations, such as the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use and the Society for Clinical Data Management, have published detailed guidance on how to design CRFs, placing their focus on ensuring accuracy, utility, and poignancy of layout and content[7],[8] , [9]. However, the idea of designing CRF variables to facilitate standardization and reusability in the context of data interoperability has remained unaddressed by these expert institutions.

In the simplest setup, a study will be limited to 1 cohort, enrolling patients within one country, across several participating medical offices or hospitals. In more complex cases, patients are enrolled across several cohorts located in multiple countries. Regardless of the complexity of the study, interoperability and quality of data should always be considered high-priority objectives when designing the collection forms. By so doing, study results are more reliable, and collected data are ready for potential secondary use.

The COVID-19 pandemic, along with disease outbreaks caused by other viruses such as human monkeypox, Zika, and Ebola, has highlighted the need for international collaboration in terms of research. They have led to large-scale clinical studies conducted in the private sector and by public research consortia [10]. Multi-country and multi-cohort retrospective studies come with more challenges, as they generally need to combine data that were collected in different formats and, at times, different languages. This means that even when variables cover similar information, extensive transformation or translation activities are required before merging data can take place. This highlights the need for coherent study protocols across research groups and across countries, based on common formats and terminologies. Here is where data standardization and harmonization are the key to enabling quality data that can be merged easily without resource-heavy transformation activities, thus expediting analysis and gaining timely insights [11,12].

Aim

Our aim in this viewpoint is to provide researchers with general suggestions in designing CRFs that, in our opinion, should support interoperability, reduce ambiguity, improve data quality, and facilitate data exchange across different systems.

The ORCHESTRA Project

ORCHESTRA was funded by the European Commission during the COVID-19 pandemic and ran until the end of 2024. It aimed at creating a pan-European cohort of COVID-19 patients to study the disease, the efficacy of the treatment, and the long-term effects on general and fragile population as well as on health care workers. We were responsible for establishing interoperability within the European ORCHESTRA project. Partners in ORCHESTRA followed the goal to merge data from different clinical studies to generate new knowledge about multiple aspects of COVID-19 [3]. We examined over 3700 variables (comprised of questions and answers) with the objective of identifying similar information across studies that could be matched and analyzed jointly. This way, we compared the variety of approaches used across 7 different COVID-19 studies to investigate similar health care concepts. Within our project task, we associated international identifying codes from the most pertinent standard health terminologies to the questions and answers included in 7 clinical studies in ORCHESTRA. Ambiguous or complex wording found in the CRF variables needed to be evaluated as part of this mapping activity as well.

Based on our experience of being involved in large-scale national and international research projects, we believe that apart from ensuring accuracy and quality of collected data, CRF design should also maximize semantic interoperability. That way, time- and cost-efficient merging, analysis, and sharing of data can be facilitated. Our conviction is supported by the report published by the Joint Action Towards the European Health Data Space, which is a European initiative that developed principles for the secondary use of health data [13] and that places semantic interoperability as one of the operational objectives to achieve excellence of data quality.

To address the pressing need for streamlined data exchange and integration in clinical research, we have formulated 5 guiding principles that should be considered when designing CRFs. The principles address the need to harmonize data, unambiguously identify variables, associate clinical concepts with international identifiers, promote data quality, and enhance semantic coherence.

We believe that the application of the proposed principles would enhance semantic interoperability and support the exchange of information across different research groups.

Five Guiding Principles to Enhance Interoperability of CRFs

Following our aim statement, we propose 5 guiding concepts aimed at increasing semantic interoperability of CRF data and quality of collected information.

XSL•FO
RenderX

Harmonize Data

Concept 1: Creation or Reuse of Core Data Elements (CDEs)

Recurring information across clinical trials that are general in nature or specific to a specific disease should be identified and shared in a common format across the scientific community.

Data concerning demographics or clinical evaluation of patients, for example, is collected in many studies and should ideally be standardized to create a uniform format. Frequently collected disease progression and outcome information should also be identified and grouped into well-defined disease-specific core data elements (CDEs). CDEs can then be published for re-use by researchers worldwide.

Granted, the National Institutes of Health (NIH) have been focusing on developing CDEs for over 20 years. Yet, the adoption of CDEs is facing challenges. Reasons for low adoption by research projects, among others, are established data collection practices and ambiguous interpretations and implementations of health care concepts and CDEs.

The latter difficulty can, however be overcome by associating a common standard terminology to data elements to remove any ambiguity of meaning.

The increasing need to combine data in order to address global threats to life will eventually have to gain greater weight than maintaining localized, established practices.

National and international harmonization efforts such as the International Patient Summary (IPS) [14], the European Electronic Health Record Exchange Format [15], or the United States Core Data for Interoperability (USCDI) [16] should be considered to maximize reusability of data.

Unambiguously Identify Variables

Concept 2: Use of Standards in CRF Metadata

Following the FAIR principles, metadata is of foremost importance for the quality re-use of information. Metadata should ideally include references to international terminology codes that unambiguously represent each concept. Whenever possible, all CRF variables should be handled as close-ended questions. In case of measurement or observations, Logical Observation Identifiers Names and Codes (LOINC) should be used. LOINC offers the possibility of defining data elements without ambiguity by specifying with one code all the details relative to the observation to be performed.

When no LOINC codes are available, other international terminologies should be considered, such as the National Cancer Institute's Thesaurus for genomics data.

Qualitative answers should be restricted to defined (coded) value sets and identified with the appropriate terminology codes, such as those provided by Systematized Medical Nomenclature for Medicine–Clinical Terminology (SNOMED CT) [17]. The Anatomical Therapeutic Chemical Classification System can be used to describe drugs and chemicals, and the *International Classification of Diseases* to report diseases and disorders.

CRFs often include more complex questions that cover several informational components. That can include details about time, location, situation, etc. Hence, mapping several concepts to one semantic code can at times be difficult. This is a challenge that the postcoordination expression in SNOMED CT may help to solve in some contexts. However, integrating postcoordination into CRFs appears less feasible [18].

Terminology bindings proposed by national or international harmonization initiatives such as those mentioned in Concept 1 should be reflected in the metadata of the study data elements whenever possible. It should also be noted that the metadata could also include information on the format of the data. That can be achieved by mapping study elements to standards such as the Health Level Seven's (HL7) Fast Healthcare Interoperability Resources (FHIR) standard or to the Observational Medical Outcomes Partnership Common Data Model.

The combined use of semantic and syntactic standards would further support interoperability.

Associate Clinical Concepts With International Identifiers

Concept 3: One Concept at a Time (Unless It Is Part of a Questionnaire or Index)

The progressive adoption of FHIR by initiatives aiming to harmonize health data, such as IPS and USCDI, would suggest that information should follow the modular structure and be as precise as possible.

Hence, our second recommendation would be to only include one concept in a question. We have, for example, seen the following question in a COVID-19–related study CRF: asking enrolled patients whether they have had "Changes in or a loss of smell." In this case, we propose splitting the question into 2 variables: "Changes in smell" and "Loss of smell." With this approach, each variable could be represented by a specific semantic code. Additionally, splitting the question into 2 would also facilitate accurate analysis.

We acknowledge that this accuracy has to be balanced against the manageability of the length of a CRF. It should not lead to the creation of overly long CRFs, but rather to a focus on accuracy and key variables needed for analysis.

However, it is also important to note that at times some concepts which are included in questions could be removed without lowering accuracy of the question. For example, information like time points or target patients could be put as a header or as instructions within the CRF.

Of course, the ultimate decision on what constitutes the most relevant concepts to be included in the CRF questions always lies with the scientific group or the principal investigator or the sponsor.

Promote Data Quality

Concept 4: Accurate Wording

Initiatives such as IPS and the USCDI provide guidance on the use of patient-related core data elements. However, often CRFs

XSL•FO
RenderX

require the use of more study-specific variables. The wording of variables (questions and answers) should be carefully phrased to provide all and only the necessary information so that study nurses and respondents understand exactly what is asked. Precisely worded questions (and answer options) will increase the quality of data. This is to ensure that the data collected in response to these variables will be comparable and ready to be merged. In the case of laboratory examinations, the use of LOINC codes can be very useful because it automatically includes all the information necessary to remove ambiguity, that is, the methodology used, specimen type, or whether the expected result is qualitative or quantitative, unit information, etc.

If language translations of CRF variables are required, these should be meticulously performed. Additionally, a quality check on the wording and meaning of the translated variables is recommended and should be implemented, ideally by native speakers.

Enhance Semantic Coherence

Concept 5: Answer Options Should All Be Semantically Coherent, and Units of Measures Clearly Stated and Identified With the Unified Code for Units of Measure Units

Another aspect of standardization relates to the answer options that are the second component of a CRF variable, after the question. It is important to maintain coherent semantic coding of answers as well. We recommend that, during the design of CRFs, semantic codes be used for mapping answers instead of assigning generic numeric identifiers that lack specificity. This would provide an unambiguous and reusable representation of the answer concept. Furthermore, coding would help the precision of the information and the quality of data by highlighting inconsistencies such as value set options not being semantically aligned.

For example, the question “What kind of swab test was performed?” should not include “throat” and “PCR” as answers in the same value set. In this case, the use of a terminology system like SNOMED CT clearly shows that the codes of the 2 concepts belong to different semantic categories (“body structure” and “procedure” respectively). Therefore, the question might be equivocal and could lead to unclear results.

In addition, in case of variables that describe quantitative (laboratory) measurements, such as “Dose of immunosuppressive medication taken per day” “Body weight” or “Glucose concentration measured” units of measure should be clearly stated and identified with Unified Code for Units of Measure codes [19,20].

Summary and Considerations

Based on our experience in standardizing and harmonizing CRF variables from different protocols on COVID-19, we have presented five concepts aimed at improving CRF design and enhancing interoperability of clinical study data: (1) the creation of or (2) use of already existing standardized data objects can save time and help establish alignment and comparability with other research datasets, and (3) data quality can significantly be improved by paying attention to the fact that each CRF question and every answer option only contains 1 concept, (4) that variables are accurately phrased, and (5) that answer options are coherent, or in case of numerical results, that clearly defined units are included.

Probably, guidelines for reporting multiple concept codes in the metadata should be established, as this is a common occurrence in clinical research. In many cases, the possibility to code complex questions as coded questionnaires is very helpful. We believe it would be important to collaboratively address this difficulty of coding complex variables for the clinical study context.

The problem of not using interoperability standards is invisible to many researchers. That is because often, the advantages of such use become evident only when the need to merge data arises. Ideally, standards should be introduced already during the design phase of CRFs. Unfortunately, since the importance of data standards is still not adequately known, their implementation might be seen by some as a hassle that slows down or even limits the development of CRFs. In our opinion, a cultural change based on education and information in this field is needed.

It is necessary to abandon the idea of doing research in silos with data collected in incompatible formats by different research groups. Common data elements and their format should be identified, agreed on, and promoted by relevant national and international authorities. On the other hand, a scenario where clinicians are unwillingly responsible for standardization should be avoided, considering the already existing strain on their time and understandable gaps in expertise. New roles in health care are needed with expertise in digital medicine to enable interoperability of data and facilitate their integration within a wider eHealth ecosystem where data are being collected from different digital solutions and in a cross-country context.

The use of a common exchange format in combination with standard terminologies for study data elements in an eCRF would complete the interoperable data model of clinical research information. The innovative standard that is increasingly being adopted in the health care environment is HL7 FHIR. Thanks to FHIR’s innovative modular organization of information, a particularly efficient exchange of data is enabled. Its adoption in the clinical research environment is still low, but we expect this to change in the future. That is because the need to streamline activities and integrate information from electronic patient records or other medical devices into clinical research is progressively becoming more evident [21,22]. In fact, a dedicated HL7 working group is focusing on the design of FHIR resources to conduct clinical research more effectively [23].

In this context, as mentioned before, the European Commission has proposed a regulation for the European Health Data Space to support the interoperability of data in healthcare and in research. Forthcoming implementing acts will provide specifications for the exchange format of data to support cross-system and cross-border portability. In the United States, the Office of the National Coordinator for Health Information Technology promotes the USCDI [16], a set of health data

XSL•FO
RenderX

elements divided into thematic classes to support information exchange. In general, collaborations to improve interoperability are fortunately increasing, including, for example, the Observational Health Data Sciences and Informatics (OHDSI) and the NIH concerning the maternal health data [24] or the employment of the Observational Medical Outcomes Partnership Common Data Model by the NIH research program "All of Us" [25]. Very important is also the OHDSI and European Medicines Agency collaboration in the project DARWIN [26], and the HL7 Vulcan project bringing together OHDSI and HL7 FHIR [27].

In conclusion, the COVID-19 pandemic has revealed to the broad community how important it is to quickly analyze large amounts of data, develop vaccines, and assess their safety and efficacy. We therefore need to facilitate the exchange of information in the context of global health challenges (including cancer, infectious disease, and rare diseases) and implement standardization of clinical study data collection. Additionally, the establishment of an international coordination board for standards and interoperable clinical study data with competence in clinical data, interoperability standards, and data protection should be part of a preparedness plan to face future pandemics and other health threats. This proposed coordination board, in coordination with ongoing international initiatives, could be instituted at the regional level and associated with large funding bodies and policy makers (ie, European Commission within the European Union and NIH in the United States) or at the pan-regional level, for example, as part of the World Health Organization.

Acknowledgments

We would like to thank Prof Evelina Tacconelli for her invaluable support. The guiding principles were conceived based on the activities performed within the ORCHESTRA project, which has received funding from the European Union's Horizon 2020 research and innovation program (grant 101016167).

Data Availability

The metadata definitions on which we based our comment are publicly available on the standard-enabling platform ART-DECOR [28].

Authors' Contributions

ER and CS contributed equally conceiving and designing the guiding principles. ST has revised the paper critically and made important contributions.

Conflicts of Interest

None declared.

References

1. Sass J, Bartschke A, Lehne M, et al. The German Corona Consensus Dataset (GECCO): a standardized dataset for COVID-19 research in university medicine and beyond. Research Square. Preprint posted online on Dec 21, 2020. [doi: 10.21203/rs.3.rs-51348/v3]
2. ORCHESTRA Cohort. URL: https://orchestra-cohort.eu/ [accessed 2024-12-08]
3. Rinaldi E, Stellmach C, Rajkumar NMR, et al. Harmonization and standardization of data for a pan-European cohort on SARS- CoV-2 pandemic. NPJ Digit Med 2022 Jun 14;5(1):75. [doi: 10.1038/s41746-022-00620-x] [Medline: 35701537]
4. Kessissoglou IA, Cosgrove SM, Abboud LA, Bogaert P, Peolsson M, Calleja N. Are EU member states ready for the European Health Data Space? Lessons learnt on the secondary use of health data from the TEHDAS Joint Action. Eur J Public Health 2024 Dec 1;34(6):1102-1108. [doi: 10.1093/eurpub/ckae160] [Medline: 39514641]
5. Solle D. Be FAIR to your data. Anal Bioanal Chem 2020 Jul;412(17):3961-3965. [doi: 10.1007/s00216-020-02526-7] [Medline: 32300841]
6. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. Perspect Clin Res 2014 Oct;5(4):159-166. [doi: 10.4103/2229-3485.140555] [Medline: 25276625]
7. Guideline for good clinical practice E6(R2). European Medicines Agency. 2016 Jan 12. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-good-clinical-practice-e6r2-step-5_en.pdf [accessed 2023-02-14]
8. Good clinical data management practices. Society for Clinical Data Management. 2023 Oct. URL: https://scdm.org/wp-content/uploads/2019/10/21117-Full-GCDMP-Oct-2013.pdf [accessed 2023-03-20]
9. Richesson RL, Nadkarni P. Data standards for clinical research data collection forms: current status and challenges. J Am Med Inform Assoc 2011 May 1;18(3):341-346. [doi: 10.1136/amiajnl-2011-000107] [Medline: 21486890]
10. Beall RF, Moradpour J, Hollis A. The private versus public contribution to the biomedical literature during the COVID-19, Ebola, H1N1, and Zika public health emergencies. PLoS One 2021;16(10):e0258013. [doi: 10.1371/journal.pone.0258013] [Medline: 34679120]

XSL•FO
RenderX

11. Lin CH, Wu NY, Liou DM. A multi-technique approach to bridge electronic case report form design and data standard adoption. J Biomed Inform 2015 Feb;53:49-57. [doi: 10.1016/j.jbi.2014.08.013] [Medline: 25200473]
12. Fegan G, Cheah PY, Data Sharing Working Group. Electronic address: WorkingGroups@covid19crc.org. Solutions to COVID-19 data sharing. Lancet Digit Health 2021 Jan;3(1):e6. [doi: 10.1016/S2589-7500(20)30273-9] [Medline: 33735070]
13. TEHDAS. URL: https://tehdas.eu/ [accessed 2025-02-24]
14. Kay S, Cangioli G, Nusbaum M. The international patient summary standard and the extensibility requirement. Stud Health Technol Inform 2020 Sep 4;273:54-62. [doi: 10.3233/SHTI200615] [Medline: 33087592]
15. COMMISSION RECOMMENDATION (EU) 2019/ 243 - of 6 February 2019 - on a European Electronic Health Record exchange format. : Official Journal of the European Union; 2019 URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019H0243#:~:text=(1)%20This%20Recommendation%20sets%20out,health%20data%20in%20the%20Union [accessed 2025-02-24]
16. United States Core Data for Interoperability (USCDI). Office of the National Coordinator for Health Information Technology. URL: http://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi [accessed 2024-05-15]
17. Bietenbeck A, Streichert T. Preparing laboratories for interconnected health care. Diagnostics (Basel) 2021 Aug 17;11(8):1487. [doi: 10.3390/diagnostics11081487] [Medline: 34441421]
18. Cornet R, Nyström M, Karlsson D. User-directed coordination in SNOMED CT. Stud Health Technol Inform 2013;192:72-76. [Medline: 23920518]
19. Unified Code for Units of Measure (UCUM). National Library of Medicine. URL: https://ucum.nlm.nih.gov/ [accessed 2023-03-20]
20. Rajput AM, Ballout S, Drenkhahn C. Standardizing the unit of measurements in LOINC-coded laboratory tests can significantly improve semantic interoperability. Stud Health Technol Inform 2020 Nov 23;275:234-235. [doi: 10.3233/SHTI200733] [Medline: 33227779]
21. Pétavy F, et al. Global standardization of clinical research data. Appl Clin Trials 2019;28(4):2023 [FREE Full text]
22. Kush RD, Nordo AH. In: Richesson RL, Andrews JE, editors. Data Sharing and Reuse of Health Data for Research: Springer International Publishing; 2019:379-401. [doi: 10.1007/978-3-319-98779-8_18]
23. Leroux H, Denney CK, Hastak S, Glover H. A framework for representing clinical research in FHIR. Presented at: 12th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences; Dec 10-11, 2019; Edinburgh, Scotland URL: https://ceur-ws.org/Vol-2849/paper-04.pdf [accessed 2025-02-24]
24. Johns Hopkins awarded NIH maternal health research data innovation and coordination hub. Biomedical Informatics and Data Science (BIDS). URL: http://dhsi.med.jhmi.edu/news/johns-hopkins-awarded-nih-maternal-health-research-data-innovation-and-coordination-hub [accessed 2024-05-16]
25. Ronquillo JG, Lester WT. Precision medicine landscape of genomic testing for patients with cancer in the National Institutes of Health All of Us Database Using Informatics Approaches. JCO Clin Cancer Inform 2022 Jan;6:e2100152. [doi: 10.1200/CCI.21.00152] [Medline: 34985965]
26. Dernie F, Corby G, Robinson A, et al. Standardised and reproducible phenotyping using distributed analytics and tools in the Data Analysis and Real World Interrogation Network (DARWIN EU). Pharmacoepidemiol Drug Saf 2024. [doi: 10.1002/pds.70042]
27. FHIR to OMOP FHIR IG. URL: https://build.fhir.org/ig/HL7/fhir-omop-ig/ [accessed 2024-05-16]
28. Orchestra - project information. ART-DECOR. URL: https://art-decor.org/art-decor/decor-project--orch [accessed 2025-02-24]

Abbreviations

CDE: core data element
CRF: case report form
eCRF: electronic case report form
FHIR: Fast Healthcare Interoperability Resources
HL7: Health Level Seven
IPS: International Patient Summary
NIH: National Institutes of Health
OHDSI: Observational Health Data Sciences and Informatics
SNOMED CT: Systematized Medical Nomenclature for Medicine–Clinical Terminology
USCDI: United States Core Data for Interoperability

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 04.08.23; peer-reviewed by B Senst, B Feng, C Asuzu, F Amar; revised version received 08.12.24; accepted 08.01.25; published 03.03.25.

Please cite as:
Rinaldi E, Stellmach C, Thun S
How to Design Electronic Case Report Form (eCRF) Questions to Maximize Semantic Interoperability in Clinical Research
Interact J Med Res 2025;14:e51598
URL: https://www.i-jmr.org/2025/1/e51598
doi:10.2196/51598

© Eugenia Rinaldi, Caroline Stellmach, Sylvia Thun. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 3.3.2025. 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

The Evolution of Uroflowmetry and Bladder Diary and the Emerging Trend of Using Home Devices From Hospital to Home

Ming-wei Li[1]; Yao-Chou Tsai[1,2]; Stephen Shei-Dei Yang[1,2]; Yuan-Hung Pong[3]; Yu-Ting Tsai[4]; Vincent Fang-Sheng Tsai[1,5]

[1]Division of Urology, Department of Surgery, Taipei Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, New Taipei City, Taiwan

[2]School of Medicine, Tzu Chi University, Hualien, Taiwan

[3]Department of Urology, Ten Chen Hospital, Taoyuan City, Taiwan

[4]PhD Program of Mechanical and Aeronautical Engineering, Feng Chia University, Taichung, Taiwan

[5]Department of Urology, National Taiwan University Hospital, Taipei city, Taiwan

Corresponding Author:
Vincent Fang-Sheng Tsai
Division of Urology, Department of Surgery
Taipei Tzu Chi Hospital
Buddhist Tzu Chi Medical Foundation
16F, No 289, Jianguo Road
New Taipei City, 231016
Taiwan
Phone: 886 952 691807
Email: ntubala@gmail.com

Abstract

Although uroflowmetry and bladder diaries are widely used for noninvasive evaluation of lower urinary tract symptoms, they still have limitations in diagnostic capability and users' convenience. The aim of this paper is to discuss potential solutions by reviewing (1) the evolution and current clinical use of uroflowmetry and bladder diary, including clinical guidelines, daily practice applications, and their historical development; (2) a growing trend toward using home devices with various technologies; and (3) a comprehensive comparison of the strengths and weaknesses of these home devices. In our opinion, the following points can be highlighted: (1) the emerging trend of using home devices can enhance diagnostic capabilities through repeated measurements and the convenience of at-home testing and (2) home devices, which provide both frequency-volume and uroflowmetry information, have the potential to transform the management of lower urinary tract symptoms.

(Interact J Med Res 2025;14:e66694) doi:10.2196/66694

KEYWORDS

lower urinary tract symptoms; uroflowmetry; bladder diary; home devices; bladder; noninvasive; evaluations; viewpoint; diagnostic; mobile health

Introduction

As the population ages, there has been an increase in patients reporting lower urinary tract symptoms (LUTS) in recent years. To objectively assess the function of the lower urinary tract, uroflowmetry and bladder diaries (BD) are commonly used noninvasive examinations for those experiencing LUTS.

Uroflowmetry can measure various parameters during the voiding phase, including maximum flow rate (Qmax), and voided volume, which are essential metrics. Additional parameters, such as flow pattern, time to Qmax, flow time, and average flow rate, provide further insights [1]. This makes uroflowmetry a comprehensive and objective tool for evaluating the voiding phase. Consequently, uroflowmetry is recommended in the guidelines of both the American Urological Association and the European Association of Urology for assessing LUTS [2-4].

BD offers physicians valuable insights into urinary frequency, functional bladder capacity, most frequently observed bladder volume, urgency and urge urinary incontinence episodes, volume of water intake, nocturnal or daily urine output ratio, and associated bladder pain episodes [5]. They assess not only the storage phase but also some symptoms related to the voiding phase. For optimal adherence and reliability, a duration of 3-7 days is recommended for maintaining a BD [3,6]. Additionally,

XSL•FO
RenderX

BD is included in the guidelines of both the American Urological Association and the European Association of Urology for evaluating overactive bladder and LUTS [2-4].

Even though both of them can provide physicians with several parameters for evaluating the lower urinary tract, there are still some limitations of uroflowmetry and BD for depicting the whole picture of patients' LUTS, such as inconveniences, inadequate measurement frequency for uroflowmetry, lack of objective recording, and poor adherence for BD.

The aim of this paper is to discuss potential solutions for the earlier-mentioned limitations by reviewing (1) the evolution and current clinical use of uroflowmetry and BD, including clinical guidelines, daily practice applications, and their historical development; (2) a growing trend toward using home devices with various technologies; and (3) a comprehensive comparison of the strengths and weaknesses of these home devices.

Clinical Application and Limitation of Uroflowmetry and BD

Clinical Application and Limitation of Uroflowmetry

Uroflowmetry is typically performed in health care institutes, where patients may experience heightened emotional effects compared to when they are at home. Pre- and intratest anxiety (just like "white-coat hypertension") should be taken into consideration for its effect on lowering test reproducibility [7]. This practice also causes inadequate measurement frequency and a lack of voiding in different scenarios.

As to representativeness, some parameters are prone to within-subject variation [8-11], so it is recommended to repeat uroflowmetry measurements. The diagnostic accuracy of uroflowmetry is largely affected by threshold values [12,13], especially with physiological compensatory processes, detrusor underactivity, or an underfilled bladder [14]. Although uroflowmetry can be used to monitor treatment outcomes [15] and to correlate symptoms with these objective findings [12,16], its clinical value is still limited, as it is unable to differentiate between the possible underlying mechanisms. Again, specificity can be improved by repeated flow rate testing [6]. However, it is not feasible for a patient to receive multiple uroflowmetry in a clinical setting. Therefore, Caffarel et al [17] and Bray et al [18] even suggested the use of home uroflowmetry to achieve multiple measurements to improve the reliability of uroflowmetry.

In summary, uroflowmetry should be performed in a comfortable environment repeatedly to represent the whole picture of a patient's voiding pattern. This means that an at-home uroflowmetry is preferred.

Clinical Application and Limitation of BD

BD provides a method of quantifying symptoms, such as frequency of urge incontinence events and number of nocturia episodes [3], and reduces recall bias. However, it is not completely objective because the process relies on manual actions, such as recording the time and using measurement tools. These will affect BD's reliability.

Furthermore, the use of BD might induce a "bladder training" effect and influence the frequency of nocturnal voids [19]. Though it is important to guide patients to live an appropriate lifestyle [20], such modifications could potentially obscure the true micturition behavior of the patient. It is just like a clinical "uncertainty principle" (when you measure a system and change it at the same time), which should be taken into consideration when the goal of the test is to accurately depict the true picture of a patient's micturition. It is preferred to record a BD intuitively and insensibly.

The duration of BD has to be long enough to reduce sampling errors but short enough to avoid nonadherence [21]. It is recommended to conduct a BD including at least 3 days (continuous or separated) [22]. However, a longer BD is generally more reliable [23-26]. Because of the considerable effort required to complete each entry in a BD, there has been minimal achievement in comparing and validating BDs. Until now, the International Consultation on Incontinence Questionnaire BD is the only one that has undergone full validation [27]. The current protocol of BD necessitates that patients physically collect and measure each void, recording the volume along with associated symptoms. This practice requires patients' thorough intellectual understanding of the whole procedure and objectively recording BD. Among a survey of urogynecology or female pelvic medicine and reconstructive surgery specialists' fellowship-trained attendings, 25% had never or rarely (frequency of use <20%) used BD, and 97.5% reported difficulties associated with obtaining correctly completed and clinically applicable BD. The authors mentioned the best way is to teach patients with thorough instruction [20]. Therefore, a BD system capable of automation, objective recording, and repeated measurement appears to be the ideal solution.

Innovative and Possible Solutions for Avoiding Limitations

Evolution History of Uroflowmetry

As discussed in the "Clinical Application and Limitation of Uroflowmetry" section, it is important to provide patients with a comfortable environment and repeated measurements. In a clinical setting, urine weight–based, dipstick, and spinning-disk uroflowmetry are currently the most common conventional measurements of flow rate. Urine weight–based uroflowmetry detects the rate of changed weight or volume of voided urine. Dipstick uroflowmetry uses a dipstick immersed in voided urine sensing the change of fluid height and obtaining the flow rate. Spinning-disk uroflowmetry relates the power needed to counteract the slowing speed of the spinning disk hit by falling urine to the rate of urine flow [28]. In this paper, some trials of home uroflowmetry (relatively comfortable environment and feasible for repeated measurements) will be presented and discussed.

There have been several types of funnel-based uroflowmetry in the market or research fields [29-36]. The principle of funnel-based uroflowmetry is a combination of a funnel and some calibrating assembly for measuring Qmax. They all need

XSL•FO
RenderX

to be operated manually, and the results are then recorded on paper. One funnel-based home uroflowmetry has been developed for decades [37]. P-Flow (Tejnaksh Healthcare Ltd) is now a commercially available home uroflowmetry, which provides not only Qmax, average flow, voiding time of urine, flow curve, and total voided volume of urine but also simple urinalysis by an accompanied dipstick. However, users need to upload photos of the indicators on the device after finishing the test onto the health care company's server for interpreting the final results. There would be a time lag before obtaining the interpretation from the experts of this company. Additionally, it was found that Tejnaksh, the company behind this product, is a recognized teaching institute by the Ministry of Health and Family Welfare in India. To our knowledge, however, the product itself has not obtained certification from any other health authority.

There are at least 2 series of active research groups devoted to sound-based uroflowmetry, one from Singapore [38] and another from Korea [39]. Both systems use mobile phones to detect voiding sound and transfer it to the signal of voiding. Important urodynamic parameters, such as Qmax, voided volume, voiding time and average flow rate, and voiding flow patterns, can be presented through the embedded artificial intelligence (AI) algorithm. ProudP is currently commercially operated in the United States, which not only enhances the communication between patients and doctors but also improves the care quality for patients with LUTS. However, sound sensing uroflowmetry may be interfered in some noisy environments and even by the material of the testing urinal. There will be more discussion about the features of home devices in the "Comparison Between Home Devices of Different Technologies" section.

To provide a comfortable and intuitive environment for uroflowmetry, a toilet-shaped uroflowmetry (UM-100, Toto Ltd) was commercially developed in 2008 and was initially designed for clinical use [40]. A water-level measurement unit is installed with a connection to the toilet bowl. By balancing the hydraulic pressure between the measurement unit and the toilet bowl (communicating tubes in physics), urination-caused change in the water level can be measured and transformed to uroflowmetry parameters. The merit of this system is no necessity to do any cleaning work after each measurement. However, since the system detects the change in the total volume, it may have erroneous measurements when patients pass stool and urinate simultaneously [41].

A novel technology of vibration-based uroflowmetry was developed by a Taiwan-based interdisciplinary team [42]. This system simultaneously detects vibration signals during urination using an accelerometer alongside conventional uroflowmetry. Strong correlations were observed between this system and conventional uroflowmetry for parameters such as Qmax, voided volume, voiding time, and time to Qmax. Additionally, an AI model was used to analyze and predict 6 predefined patterns of uroflow curves, aiding in diagnosing voiding dysfunction with an accuracy of approximately 98%. This relatively low-cost system is suitable for automatic home urinary monitoring and enables repeated uroflow monitoring of patients outside health care institutions.

Evolution History of BD

With the reflection in the "Clinical Application and Limitation of BD" section, it is suggested to provide patients with automation, objective recording, and repeated measurements. It can be traced back to 1993 when the first computerized voiding diary "Compu-Void" was developed [43]. Compu-Void was a 64,000 RAM–capacity manual unit operated on the primitive operating system "DOS" of a personal computer. In the following decades, numerous trials have been conducted in pursuit of enhancing BD, aiming for automation, objective recording, and repeated measurement.

Quinn et al [44] ever developed a logical flow for asking users' symptoms on the portable electronic BD "MiniDoc." The advantage of an electronic data-inputting device lies in its accuracy and speed in retrieving data for further analysis [44]. Mangera et al [45] compared a paper-card reader and a hand-held input device with conventional written BD and found that an intuitive and user-friendly interface led to not only patients' preference but also the efficiency and accuracy of data management. With the onset of the digital era, application software (apps) on mobile phones and tablet computers began to play a crucial role in recording BD [46], and their numbers surged rapidly. Reports indicated that there were 55 apps available in languages such as Portuguese, Spanish, French, or English, with some offering functions for analyzing incontinence episodes and nocturia [47]. However, all of these apps require manual input.

For automated entry of data, Takai et al [48] ever introduced a body weight–based BD, which demonstrated a strong correlation between the voided urine weight recorded by the device and voided urine weight measured manually by the examinee [48]. This is the best practice of frequency or volume record and should be combined with entering episodes of LUTS to accomplish a comprehensive BD, though it cannot measure accurate urine weight during defecation.

With robust procedures of BD, the most common way to teach patients how to complete a BD is by providing detailed instructions along with any type of BD. In the current digital age, it should be expected that more clinicians like to use mobile apps or other digital resources. However, it was shown that very few clinicians actually used mobile apps (0.9%) or directed patients to use internet resources (1.2%) [20]. Furthermore, there is still no evidence to suggest that these digital apps can independently improve return rates or accuracy of the BD. In our opinion, the reason behind this is that there is only little difference between paper-based and electronic-based BDs. Electronic BDs simply record data through manual input or touch-screen processes, lacking automation. Moreover, the most cumbersome step of BD, collecting and measuring urine volume, remains unchanged. In this regard, a body weight–based BD offers certain advantages [48].

Since some important characteristics of a reliable BD, such as automation, objective recording, and repeated measurement, are fulfilled by the earlier-mentioned technologies (Table 1), we can envision a more advanced tool for accurately depicting a patient's voiding patterns. Additional suggestions for the development of a combined home uroflowmetry and BD system

XSL•FO
RenderX

are as follows: elimination of the need for cleaning after each use, incorporation of an intuitive and user-friendly interface, integration of wireless communication capability, accessibility of data for both users and doctors, and suitability for both storage and voiding phases. Therefore, such a comprehensive and intuitive "voiding recorder" will be a new standard for at-home voiding monitoring.

Table 1. Chronology of uroflowmetry and bladder diary (BD) advanced technologies in the recent century.

Year	Event	Remarks
1932	Ballenger suggested the maximum distance of a man's urine ejecting	[49]
1948	Drake recorded change in urine weight with time and manually calculated flow rate	[49,50]
1957	Kaufman improved Drake's system with electrical apparatus	[51]
1957	Von Garrelts electrically calculated flow rate	[41]
1965	Smith designed funnel-based uroflowmetry	[29]
1976	Drach used a dipstick to estimate the flow rate	[35]
1993	The first computerized voiding diary "Compu-Void" was developed	[43]
1999	International Continence Society defined uroflowmetry parameters	[52]
2003	Quinn drew a logical flow for asking users' symptoms	[44]
2008	Toto developed communicating tube uroflowmetry	[40]
2014	Mangera used paper-card reader	[45]
2014	Bright designed the first standardized and validated BD (ICIQ[a] BD)	[27]
2015	Krhut compared uroflowmetry and sonouroflowmetry	[53]
2016	Application of mobile phones and tablet computers in recording BD	[46]
2021	Takai introduced a body weight–based BD	[48]
2022	Pong linked vibration with uroflowmetry	[42]

[a]ICIQ: International Consultation on Incontinence Questionnaire.

The Emerging Trend of Using Home Devices

As we enter the era of AI, big data become crucial for training and applying AI models. This enhances the value of home devices, particularly for their ability to repeatedly measure home uroflowmetry and generate big data. Home devices facilitate mobile health or medicine by offering a comfortable environment, enabling repeated measurements, and providing big data for communication and application. It has been noted that extensive at-home data are often more reliable than single in-office tests [54]. Recently, there has been a surge in interest in home uroflowmetry at international scientific meetings and in literature [9,55-58]. Several commercial home uroflowmetry devices have also been introduced, offering repeated measurements of voiding parameters and uroflow curves. This method of repeated home measurement can provide a comprehensive picture of patients' daily-life voiding without the anxiety and stress associated with office uroflowmetry tests [9,55]. As mentioned in the "Introduction" section, uroflowmetry primarily describes the condition of the voiding phase, while BD addresses the storage phase and some symptoms related to voiding. Each can be applied to patients with different types of voiding dysfunction. Following advancements, home devices can now measure parameters in both the storage and voiding phases, making them suitable for almost all patients with voiding dysfunction.

Certainly, there is a need for home uroflowmetry to offer repeated measurements and convenience, addressing several clinical and practical challenges for practicing urologists [54]. With various home uroflowmetry technologies emerging in the market, it is important to understand their strengths and weaknesses. What are the comparative advantages and disadvantages of these different home uroflowmetry technologies?

Comparison Between Home Devices of Different Technologies

All types of home uroflowmetry technologies share common strengths, such as automatic recording, repeated measurements, and the generation of big data. However, each technology also has its own unique strengths and weaknesses when used for home uroflowmetry. Table 2 outlines various features of these technologies, including accuracy, susceptibility to environmental interferences, uroflow curve pattern recognition, AI algorithms, and contact-free operation.

XSL•FO
RenderX

Table 2. The comparison on different technologies for home uroflowmetry.

Features or technologies	Weighing (gravimetric)	Height sensor	Sound	Vibration
Accuracy	FDA[a] certified [57]	FDA certified [59]	Correlation with office uroflowmetry (R=0.91) [54]; prediction rate of 99% [60]	Uroflow curve pattern recognition accuracy>0.98 [42]
Vulnerability to the surrounding interferences	No	No	Yes [61]	No
Uroflow curve pattern recognition	No	No	Yes [60]	Yes
AI[b] algorithm or model	No	No	Yes [60,61]	Yes
Contact-free (no need for installation or cleaning)	No	No	Yes	Yes

[a]FDA: Food and Drug Administration.

[b]AI: artificial intelligence.

In terms of accuracy, all technologies yield results comparable to office uroflowmetry, with some even receiving Food and Drug Administration approval. However, sound-based technologies are sensitive to environmental interferences like noise or barriers due to their measuring mechanism. For accurate recording, the urine stream must be voided into a water-filled commode rather than a urinal [54]. Additionally, the sound of voiding varies between men and women, likely due to anatomical and postural differences during urination. Men, who typically urinate standing up, produce louder sounds, as there are no barriers to dampen the sound. In contrast, women, who usually urinate sitting down, produce quieter sounds, and the sitting position can block sound transmission. To address these issues, different models have been developed for both genders [61].

Vibration-based technology offers a potential solution by reducing the barriers to sound transmission. Vibration is transmitted more directly than sound, which travels through the air, while vibration is conveyed through the concrete toilet bowl or urinal to the sensor. Consequently, sound-based technologies are more vulnerable to surrounding noise compared to vibration-based ones [56].

After reviewing the literature, it is evident that only sound- and vibration-based technologies have used AI models or algorithms for uroflow curve pattern recognition [42,60,61]. These 2 technologies are advantageous for intuitive and contact-free measurements, as they do not require installation or cleaning for each use. The calibration of any device is crucial and must be verified during the approval process. It should be ensured that all home devices receive approval from health authorities before clinical application. Additionally, with repeated measurements using home devices, the current reference ranges for clinical urodynamic studies may be revised accordingly. However, such revisions should be undertaken only after sufficient evidence has been accumulated. These home devices not only enhance accessibility for patients and physicians in assessing voiding patterns, but they also improve the diagnosis and treatment process by collecting more comprehensive information. The use of home devices is anticipated to transform the management of LUTS.

Conclusions

Uroflowmetry and BD are key diagnostic tools for LUTS, despite facing limitations such as the absence of a comfortable environment and infrequent measurements for uroflowmetry as well as the lack of automation, objective recording, and repeated measurement for BD. Technological advancements have addressed some of these limitations.

In our opinion, the following points can be highlighted: (1) the emerging trend of using home devices can enhance diagnostic capabilities through repeated measurements and the convenience of at-home testing and (2) home devices, which provide both frequency-volume and uroflowmetry information, have the potential to transform the management of LUTS.

Acknowledgments

Microsoft Copilot was used for assisting wording and grammar polishing in the manuscript. The study received funding from Taipei Tzu-Chi Hospital, Buddhist Tzu-Chi Medical Foundation (TCRD-TPE-113-66).

Authors' Contributions

VFST, YHP, and YTT were responsible for the conceptualization of the study. YCT, SSDY, and YTT provided the necessary resources and supervised the project. VFST and YCT conducted the validation. VFST and MWL wrote the original draft. YCT, SSDY, YHP, YTT, and VFST reviewed and edited the manuscript.

Conflicts of Interest

None declared.

XSL•FO
RenderX

References

1. Rosier PFWM, Schaefer W, Lose G, Goldman HB, Guralnick M, Eustice S, et al. International Continence Society Good Urodynamic Practices and Terms 2016: urodynamics, uroflowmetry, cystometry, and pressure-flow study. Neurourol Urodyn 2017;36(5):1243-1260. [doi: 10.1002/nau.23124] [Medline: 27917521]
2. Sandhu JS, Bixler BR, Dahm P, Goueli R, Kirkby E, Stoffel JT, et al. Management of lower urinary tract symptoms attributed to benign prostatic hyperplasia (BPH): AUA Guideline amendment 2023. J Urol 2024;211(1):11-19. [doi: 10.1097/ju.0000000000003698]
3. Non-neurogenic female LUTS. European Association of Urology. 2024. URL: https://uroweb.org/guidelines/non-neurogenic-female-luts/chapter/citation-information [accessed 2024-12-28]
4. Management of non-neurogenic male LUTS. European Association of Urology. 2024. URL: https://uroweb.org/guidelines/management-of-non-neurogenic-male-luts/chapter/citation-information [accessed 2024-12-28]
5. Kuo HC. Interpreting the voiding diary of patients with lower urinary tract symptoms. Incont Pelvic Floor Dysfunct 2010;4(4):105-110.
6. ACOG. Practice Bulletin No. 155: urinary incontinence in women. Obstet Gynecol 2015;126(5):e66-e81. [doi: 10.1097/aog.0000000000001148]
7. Rubilotta E, Balzarro M, Castellani D, Tiso L, Panunzio A, Pirola GM, et al. The role of emotional condition in patients with lower urinary tract symptoms performing uroflowmetry. Urology 2021;148:37-46. [doi: 10.1016/j.urology.2020.10.045] [Medline: 33217454]
8. Golomb J, Lindner A, Siegel Y, Korczak D. Variability and circadian changes in home uroflowmetry in patients with benign prostatic hyperplasia compared to normal controls. J Urol 1992;147(4):1044-1047. [doi: 10.1016/s0022-5347(17)37462-1] [Medline: 1372659]
9. Summers SJ, Armstrong JM, Kaplan SA, Te AE, Le A, Heiner SM, et al. Male voiding behavior: insight from 19,824 at-home uroflow profiles. J Urol 2021;205(4):1126-1132. [doi: 10.1097/ju.0000000000001504]
10. Jørgensen JB, Jensen KM, Mogensen P. Age-related variation in urinary flow variables and flow curve patterns in elderly males. Br J Urol 1992;69(3):265-271. [doi: 10.1111/j.1464-410x.1992.tb15526.x] [Medline: 1373664]
11. Kranse R, van Mastrigt R. Causes for variability in repeated pressure-flow measurements. Urology 2003;61(5):930-934. [doi: 10.1016/s0090-4295(03)00010-4] [Medline: 12736007]
12. Reynard JM, Yang O, Donovan JL, Peters TL, Schafer W, de la Rosette JJ, et al. The ICS-'BPH' Study: uroflowmetry, lower urinary tract symptoms and bladder outlet obstruction. Br J Urol 1998;82(5):619-623. [doi: 10.1046/j.1464-410x.1998.00813.x] [Medline: 9839573]
13. Idzenga T, Pel JJ, van Mastrigt R. Accuracy of maximum flow rate for diagnosing bladder outlet obstruction can be estimated from the ICS nomogram. Neurourol Urodyn 2008;27(1):97-98. [doi: 10.1002/nau.20480] [Medline: 17600368]
14. Siroky MB, Olsson CA, Krane RJ. The flow rate nomogram: I. Development. J Urol 1979;122(5):665-668. [doi: 10.1016/s0022-5347(17)56550-7] [Medline: 159366]
15. Siroky MB, Olsson CA, Krane RJ. The flow rate nomogram: II. Clinical correlation. J Urol 1980;123(2):208-210. [doi: 10.1016/s0022-5347(17)55859-0] [Medline: 7354519]
16. Reynard JM, Peters TJ, Lim C, Abrams P. The value of multiple free-flow studies in men with lower urinary tract symptoms. Br J Urol 1996;77(6):813-818. [doi: 10.1046/j.1464-410x.1996.00097.x] [Medline: 8705213]
17. Caffarel J, Robson W, Pickard R, Griffiths C, Drinnan M. Flow measurements: can several "wrongs" make a "right"? Neurourol Urodyn 2007;26(4):474-480. [doi: 10.1002/nau.20383] [Medline: 17245778]
18. Bray A, Griffiths C, Drinnan M, Pickard R. Methods and value of home uroflowmetry in the assessment of men with lower urinary tract symptoms: a literature review. Neurourol Urodyn 2012;31(1):7-12. [doi: 10.1002/nau.21197] [Medline: 22038511]
19. Vaughan CP, Johnson TM, Goode PS, Redden DT, Burgio KL, Markland AD. Military exposure and urinary incontinence among American men. J Urol 2014;191(1):125-129. [doi: 10.1016/j.juro.2013.07.016] [Medline: 23871759]
20. Mehta S, Geng B, Xu X, Harmanli O. Current state of bladder diary: a survey and review of the literature. Int Urogynecol J 2023;34(4):809-823. [doi: 10.1007/s00192-022-05398-w] [Medline: 36322174]
21. Bright E, Drake MJ, Abrams P. Urinary diaries: evidence for the development and validation of diary content, format, and duration. Neurourol Urodyn 2011;30(3):348-352. [doi: 10.1002/nau.20994] [Medline: 21284023]
22. Yap TL, Cromwell DC, Emberton M. A systematic review of the reliability of frequency‐volume charts in urological research and its implications for the optimum chart duration. BJU Int 2006;99(1):9-16. [doi: 10.1111/j.1464-410x.2006.06499.x]
23. Dmochowski RR, Sanders SW, Appell RA, Nitti VW, Davila GW. Bladder‐health diaries: an assessment of 3‐day vs 7‐day entries. BJU Int 2005;96(7):1049-1054. [doi: 10.1111/j.1464-410x.2005.05785.x]
24. Schick E, Jolivet-Tremblay M, Dupont C, Bertrand PE, Tessier J. Frequency-volume chart: the minimum number of days required to obtain reliable results. Neurourol Urodyn 2003;22(2):92-96. [doi: 10.1002/nau.10079] [Medline: 12579624]
25. Groutz A, Blaivas JG, Chaikin DC, Resnick NM, Engleman K, Anzalone D, et al. Noninvasive outcome measures of urinary incontinence and lower urinary tract symptoms: a multicenter study of micturition diary and pad tests. J Urol 2000;164(3 Part 1):698-701. [doi: 10.1016/s0022-5347(05)67284-9]

XSL•FO
RenderX

26. Dixon CA, Nakib NA. Are bladder diaries helpful in management of overactive bladder? Curr Bladder Dysfunct Rep 2016;11(1):14-17. [doi: 10.1007/s11884-016-0343-x]
27. Bright E, Cotterill N, Drake M, Abrams P. Developing and validating the International Consultation on Incontinence Questionnaire bladder diary. Eur Urol 2014;66(2):294-300. [doi: 10.1016/j.eururo.2014.02.057] [Medline: 24647230]
28. Addla SK, Marri RR, Daayana SL, Irwin P. Avoid cruising on the uroflowmeter: evaluation of cruising artifact on spinning disc flowmeters in an experimental setup. Neurourol Urodyn 2010;29(7):1301-1305. [doi: 10.1002/nau.20846] [Medline: 20879000]
29. Smith JC. An individual uroflowmeter. Lancet 1965;1(7376):90. [doi: 10.1016/s0140-6736(65)91667-3] [Medline: 14234215]
30. Currie RJ. The Streamtest cup: a new uroflow device. Urology 1998;52(6):1118-1121. [doi: 10.1016/s0090-4295(98)00404-x] [Medline: 9836566]
31. Pel JJM, van Mastrigt R. Development of a low-cost flow meter to grade the maximum flow rate. Neurourol Urodyn 2002;21(1):48-54. [doi: 10.1002/nau.2112] [Medline: 11835424]
32. Pridgeon S, Harding C, Newton D, Pickard R. Clinical evaluation of a simple uroflowmeter for categorization of maximum urinary flow rate. Indian J Urol 2007;23(2):114-118 [FREE Full text] [doi: 10.4103/0970-1591.32058] [Medline: 19675784]
33. Ball AJ. The peakometer: an evaluation. Urol Int 1982;37(1):42-44. [doi: 10.1159/000280794] [Medline: 6214070]
34. Colstrup H, Andersen JT, Walter S, Rytt S. A disposable uroflowmeter for recording of maximum flow rate. Accuracy and clinical acceptability. Scand J Urol Nephrol 2010;17(3):303-306. [doi: 10.3109/00365598309182136]
35. Drach GW, Binard W. Disposable peak urinary flowmeter estimates lower urinary tract obstruction. J Urol 1976;115(2):175-179. [doi: 10.1016/s0022-5347(17)59122-3] [Medline: 129576]
36. Lucas M, Moore K, Borkowski T, Pantuck A, Heesakkers J, Piotr P. Accuracy of CaptiFlow portable home flow measuring device and its use as a diagnostic tool in the management of LUTS—a randomised cross over study. 2009 Presented at: ICS Annual Meeting; 2009; San Francisco, CA, United States URL: https://www.ics.org/2009/abstract/400
37. De La Rosette JJ, Witjes WP, Debruyne FM, Kersten PL, Wijkstra H. Improved reliability of uroflowmetry investigations: results of a portable home-based uroflowmetry study. Br J Urol 1996;78(3):385-390 [FREE Full text] [doi: 10.1046/j.1464-410x.1996.00115.x] [Medline: 8881947]
38. Lee HJ, Aslim EJ, Balamurali B, Ng LYS, Kuo TLC, Lin CMY, et al. Development and validation of a deep learning system for sound-based prediction of urinary flow. Eur Urol Focus 2023;9(1):209-215. [doi: 10.1016/j.euf.2022.06.011] [Medline: 35835694]
39. Kim H, Ye C, Jung G, Ryu H, Lee JW, Jeong SJ, et al. Validation of acoustic voided volume measure: a pilot prospective study. World J Urol 2023;41(2):509-514. [doi: 10.1007/s00345-022-04231-9] [Medline: 36550234]
40. Izumi H, Nakagawa H, Kaiho Y, Miyazato M, Kawamorita N, Arai Y. Difference of uroflow parameters between novel toilet-shaped uroflowmeter (water level sensor mechanism) and conventional uroflowmeter in urologically normal female patients. 2011 Presented at: International Continence Society Annual Meeting; August 29-September 2, 2011; Glasgow.
41. Chun KS, Kim SJ, Cho ST. Noninvasive medical tools for evaluating voiding pattern in real life. Int Neurourol J 2017;21(Suppl 1):S10-S16. [doi: 10.5213/inj.1734860.430] [Medline: 28446014]
42. Pong YH, Tsai VFS, Hsu YH, Lee CH, Wang KC, Tsai YT. Application of a deep learning neural network for voiding dysfunction diagnosis using a vibration sensor. Appl Sci 2022;12(14):7216. [doi: 10.3390/app12147216]
43. Rabin JM, McNett J, Badlani GH. Computerized voiding diary. Neurourol Urodyn 1993;12(6):541-553. [doi: 10.1002/nau.1930120604] [Medline: 8312939]
44. Quinn P, Goka J, Richardson H. Assessment of an electronic daily diary in patients with overactive bladder. BJU Int 2003;91(7):647-652. [doi: 10.1046/j.1464-410x.2003.04168.x] [Medline: 12699477]
45. Mangera A, Marzo A, Heron N, Fernando D, Hameed K, Soliman AA, et al. Development of two electronic bladder diaries: a patient and healthcare professionals pilot study. Neurourol Urodyn 2014;33(7):1101-1109. [doi: 10.1002/nau.22469] [Medline: 24000163]
46. Sussman RD, Richter LA, Tefera E, Park AJ, Sokol AI, Gutman RE, et al. Utilizing technology in assessment of lower urinary tract symptoms: a randomized trial of electronic versus paper voiding diaries. Female Pelvic Med Reconstr Surg 2016;22(4):224-228. [doi: 10.1097/SPV.0000000000000266] [Medline: 27054797]
47. Vaccari NA, da Silveira LTY, Bortolini MAT, Haddad JM, Baracat EC, Ferreira EAG. Content and functionality features of voiding diary applications for mobile devices in Brazil: a descriptive analysis. Int Urogynecol J 2020;31(12):2573-2581. [doi: 10.1007/s00192-020-04382-6] [Medline: 32592019]
48. Takai S, Matsukawa Y, Hashizume N, Gotoh M. A small pilot study to evaluate the accuracy and feasibility of a novel automated voiding diary device for recording urine output measurements. Neurourol Urodyn 2021;40(1):272-277. [doi: 10.1002/nau.24547] [Medline: 33137844]
49. Drake WM. The uroflometer: an aid to the study of the lower urinary tract. J Urol 1948;59(4):650-658. [doi: 10.1016/s0022-5347(17)69425-4]
50. Chancellor MB, Rivas DA, Mulholland S, Drake WM. The invention of the modern uroflowmeter by Willard M. Drake, Jr at Jefferson Medical College. Urology 1998;51(4):671-674. [doi: 10.1016/s0090-4295(97)00203-3] [Medline: 9586627]
51. Kaufman JJ. A new recording uroflometer: a simple automatic device for measuring voiding velocity. J Urol 1957;78(1):97-102. [doi: 10.1016/s0022-5347(17)66405-x]

XSL•FO
RenderX

52. Stöhrer M, Goepel M, Kondo A, Kramer G, Madersbacher H, Millard R, et al. The standardization of terminology in neurogenic lower urinary tract dysfunction: with suggestions for diagnostic procedures. International Continence Society Standardization Committee. Neurourol Urodyn 1999;18(2):139-158. [doi: 10.1002/(sici)1520-6777(1999)18:2<139::aid-nau9>3.0.co;2-u] [Medline: 10081953]
53. Krhut J, Gärtner M, Sýkora R, Hurtík P, Burda M, Luňáček L, et al. Comparison between uroflowmetry and sonouroflowmetry in recording of urinary flow in healthy men. Int J Urol 2015;22(8):761-765. [doi: 10.1111/iju.12796] [Medline: 25988672]
54. Schultz RE. Smartphone app for in-home uroflowmetry. Urol Pract 2022;9(6):524-530. [doi: 10.1097/upj.0000000000000338]
55. Bladt L, Kashtiara A, Platteau W, De Wachter S, De Win G. First-year experience of managing urology patients with home uroflowmetry: descriptive retrospective analysis. JMIR Form Res 2023;7:e51019 [FREE Full text] [doi: 10.2196/51019] [Medline: 37847531]
56. Kumar S, Kumar M. MP09-09 mobile app-based versus conventional uroflowmetry: is your home toilet the new uroflowmetry lab. J Urol 2024;211(5S):211-229. [doi: 10.1097/01.ju.0001008920.55771.18.09]
57. Morales EVB, Peters M, Pauwels J, Vermandel A, Wachter SD, Bladt L, et al. MP75-13 digital health solutions create opportunities for home uroflowmetry. J Urol 2024;211(5S):e1237. [doi: 10.1097/01.ju.0001008676.21744.5f.13]
58. Song SH, Park JH, Park HJ, Jeong Y, Ryu H, Lee JW, et al. MP09-12 a prospective multicenter clinical trial for efficacy of mobile uroflowmetry in treatment-naive benign prostatic hyperplasia undergoing medical therapy: interim results. J Urol 2024;211(5S):e131. [doi: 10.1097/01.ju.0001008920.55771.18.12]
59. Patel D, Ogden K, Anderl W, Mastrangelo S, Le A, Hotaling J, et al. PD50-07 non-inferiority of a novel capacitance fluid height design compared to rotating disc and weight transducer designs for home uroflowmetry applications. J Urol 2018;199(4S):e972. [doi: 10.1016/j.juro.2018.02.2335]
60. Jin J, Chung Y, Kim W, Heo Y, Jeon J, Hoh J, et al. Classification of bladder emptying patterns by LSTM neural network trained using acoustic signatures. Sensors (Basel) 2021;21(16):5328 [FREE Full text] [doi: 10.3390/s21165328] [Medline: 34450769]
61. Lee YJ, Kim MM, Song SH, Lee S. A novel mobile acoustic uroflowmetry: comparison with contemporary uroflowmetry. Int Neurourol J 2021;25(2):150-156 [FREE Full text] [doi: 10.5213/inj.2040250.125] [Medline: 33387990]

Abbreviations

AI: artificial intelligence
BD: bladder diary
LUTS: lower urinary tract symptom
Qmax: maximum flow rate

Edited by T de Azevedo Cardoso; submitted 20.09.24; peer-reviewed by A Gammie, A Rantell; comments to author 19.11.24; revised version received 24.11.24; accepted 29.11.24; published 28.01.25.

Please cite as:
Li MW, Tsai YC, Yang SSD, Pong YH, Tsai YT, Tsai VFS
The Evolution of Uroflowmetry and Bladder Diary and the Emerging Trend of Using Home Devices From Hospital to Home
Interact J Med Res 2025;14:e66694
URL: https://www.i-jmr.org/2025/1/e66694
doi:10.2196/66694
PMID:

©Ming-wei Li, Yao-Chou Tsai, Stephen Shei-Dei Yang, Yuan-Hung Pong, Yu-Ting Tsai, Vincent Fang-Sheng Tsai. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.01.2025. 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

The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations

Dimitri Roustan[1], MBBS; François Bastardot[2], MD, MSACI

[1]Emergency Medicine Department, Cliniques Universitaires Saint-Luc, Brussels, Belgium

[2]Medical Directorate, Lausanne University Hospital, Lausanne, Switzerland

Corresponding Author:
Dimitri Roustan, MBBS
Emergency Medicine Department
Cliniques Universitaires Saint-Luc
Avenue Hippocrate 10
Brussels, 1200
Belgium
Phone: 32 477063174
Email: dim.roustan@gmail.com

Abstract

Large language models (LLMs) are artificial intelligence tools that have the prospect of profoundly changing how we practice all aspects of medicine. Considering the incredible potential of LLMs in medicine and the interest of many health care stakeholders for implementation into routine practice, it is therefore essential that clinicians be aware of the basic risks associated with the use of these models. Namely, a significant risk associated with the use of LLMs is their potential to create hallucinations. Hallucinations (false information) generated by LLMs arise from a multitude of causes, including both factors related to the training dataset as well as their auto-regressive nature. The implications for clinical practice range from the generation of inaccurate diagnostic and therapeutic information to the reinforcement of flawed diagnostic reasoning pathways, as well as a lack of reliability if not used properly. To reduce this risk, we developed a general technical framework for approaching LLMs in general clinical practice, as well as for implementation on a larger institutional scale.

(Interact J Med Res 2025;14:e59823) doi:10.2196/59823

KEYWORDS

medical informatics; large language model; clinical informatics; decision-making; computer assisted; decision support techniques; decision support; decision; AI; artificial intelligence; artificial intelligence tool; LLM; electronic data system; hallucinations; false information; technical framework

Introduction to Large Language Models

The development of artificial intelligence (AI) solutions and their recent democratization have allowed the public to access various innovative tools. Notably, several large language models (LLMs) have recently surged in popularity due to significant media attention and by offering free access for registered users (eg, ChatGPT, Gemini, and Meta LLaMA).

An LLM is a type of deep learning model that is pretrained on large text datasets. They are often based on the transformer architecture [1], an innovative form of neural network that uses an encoder-decoder structure to rapidly process large blocks of text, avoiding redundancies that hampered recurrent neural networks in the past. Several popular LLMs have integrated a chatbot interface to allow users to interact directly with the model, generating appropriate, context-aware responses to a user's input in a conversational manner. This allows the user to engage in dynamic conversations that appear natural, making the technology a powerful tool for various applications across a wide range of fields.

The use of LLMs has become widespread, and medicine is no exception. LLMs have the potential of becoming a disruptive tool in medicine [2] and will certainly have a major impact on clinical practice, medical education, and research. Within this field, LLM performance has already been evaluated to take specialist board exams [3], improve communication with patients [4,5], and write drafts for scientific papers. It can also be an interesting tool for creating a dynamic learning experience [6]. Although LLMs with chatbot interfaces represent an interesting tool for many tasks in clinical medicine and research, users should be aware of one of the most significant shortcomings of these models, called "hallucinations."

XSL•FO
RenderX

In this viewpoint, we review the underlying causes of hallucinations in LLMs and examine their implications within the field of clinical medicine. We also explore current and future strategies for mitigating these limitations and present a general framework to guide clinicians in critically assessing and integrating LLMs into clinical practice.

Overview of Hallucinations

In medicine, hallucinations refer to sensory experiences that occur in the absence of corresponding external stimuli. In the field of LLMs, hallucinations refer to the generation of false or fabricated information. This signifies that the LLM will create nonfactual content to answer a user's question without clarifying whether the answer contains fabricated information. Hallucinations stem from many root causes, which we will delve into below.

First, both the quality and volume of the dataset upon which the LLM has been trained are important variables and can explain the number of hallucinations the LLM produces to some degree [7,8]. How data are collected and how the model is trained can also influence hallucination frequency [8]. Furthermore, the method through which the editor fine-tunes the model can also influence the final output.

Another major cause of hallucinations stems from the very way certain LLMs are programmed. Indeed, most LLMs are auto-regressive; the term "auto-regressive" refers to the model's ability to predict future elements of a sequence based on its previous outputs. These elements, usually one or multiple words, are termed tokens. An auto-regressive LLM aims to produce an output based on token prediction; this signifies that the model will predict the most probable next token(s) given a specific input token. In practice, it predicts the following word(s) after the sequence of words it has already given. However, it generates each next token by considering the previous one, and not the whole sequence. This means words are generated in a word-after-word fashion, without necessarily using the whole of the previously generated sentence to predict the rest of the sentence [9]. This lays the groundwork for producing hallucinations since factual accuracy is not the end goal. Rather, accuracy is inferred from a high probability of adequate token prediction based on the data in the training dataset. Since the dataset is necessarily flawed or incomplete, hallucinations can arise.

The size of the training dataset can also influence hallucination type and degree. It has been demonstrated on multiple LLMs that the larger the training dataset size, the more likely the model will be capable of recognizing its limitations and acknowledging uncertainty [10]. Furthermore, choices made by the editor will also influence output quality (ie, fine-tuning decisions, output ranking, censorship, etc).

User input is also of great importance in determining the quality of the model output. Indeed, it has been shown that user input through contextualization and inclusion of source material can also modify the number of hallucinations an LLM produces [11].

Implications for Clinical Medicine

LLMs have many potential benefits in the health care system, for both providers and patients. For simple tasks, it is highly likely that part or all of the process will be carried out with LLM tools in the foreseeable future. Efficiency will likely be improved by reducing redundant and tedious tasks (most likely administrative before clinical) [12-14], and there may even be applications for reducing diagnostic delay for difficult diagnoses [15]. Nonetheless, despite these potential benefits, hallucinations represent a major risk if unaccounted for when using LLMs [16-18]. Below, we will review some situations that have appeared apparent to us when testing LLMs.

In practice, LLMs may erroneously attribute clinical, biological, or radiological characteristics to certain diseases or conditions, depending on the way the clinician inputs data as well as the probabilistic behavior of the model. This flaw, in combination with anchoring and confirmation bias, may unknowingly lead the clinician down an erroneous diagnostic or therapeutic pathway. This can have severe consequences for the patient's health.

LLMs may also make false claims about diagnostic accuracy for diagnostic procedures. This can lead the clinician to either overestimate or underestimate the diagnostic capacity of a procedure. The consequences could be either depriving a patient of a reliable diagnostic method or, on the contrary, relying on an inadequate diagnostic method to make a statement about the disease process. More specifically in the latter case, the absence of a disease process may be wrongly inferred based on an insensitive exam, and the presence of a disease process may be improperly inferred from a nonspecific exam.

Furthermore, the LLM may suggest inadequate workups and therapeutic procedures. It is important to remember that LLMs are trained on databases that may either not encompass the data necessary to provide adequate guidance (ie, absence of medical guidelines) or contain outdated medical recommendations. Further, given the crucial importance of input data supplied by the user, the omission of a simple characteristic may cause the LLM to produce an inadequate plan of care. Moreover, the LLM may not necessarily prompt the user for additional information regarding important characteristics that could influence the plan of care; most notably, social characteristics and cultural preferences may be inadequately accounted for. Specific diagnostic or therapeutic measures proposed by the LLM may be inadequate or inappropriate, based on important parameters such as pretest probability, as well as patient preferences and prognosis. In addition, given the diversity of health care system models in countries around the world, the inherent bias introduced by the LLM's dataset can lead to recommending inadequate plans of care for a different health care model than that which the dataset contains information on.

Consistency, and thus reliability, is another issue that can appear while using LLMs to make recommendations for plans of care. Indeed, even with a consistently identical user query, the information contained within the LLMs response may vary considerably when the user renews the query. This variability is an important consideration when the clinician is contemplating

XSL•FO
RenderX

the possibility of integrating LLMs into the patient care process. Indeed, it is well accepted in the health care quality community that reducing variability to a certain degree is an essential step in increasing quality [19]; this is even more so true when some parts of the information provided by the LLM are at risk of being inadequate.

In addition, the use of LLMs in clinical practice raises a significant number of ethical concerns, which we only slightly touch upon in this viewpoint. Questions regarding overreliance on LLMs and other AI tools, as well as the legal and ethical ramifications of decision-making based upon AI input, are crucial. In the era of evidence-based medicine, LLM source material and information traceability will be essential in order to reliably inform patient care decisions. The value of the clinician's experience in nuancing the LLM's outputs will also remain critical in delivering personalized patient care.

To build upon the legal and ethical concerns related to these LLMs, it is important to keep in mind that many LLMs are developed by private companies and are not open source. Data management, especially related to patient privacy rights, is an essential concern related to information input into the LLM. Implementation of LLM components in electronic medical record programs is being considered and poses the same risks. Notably, the question stands on the use of personal data to for-profit ends (targeted advertising or selling health care data to insurance companies). Local institutional governance and national regulations will be essential in managing and mitigating these potential risks.

Finally, LLM use also carries considerable potential to alter the patient-clinician relationship. Patients may increasingly discuss their symptoms and conditions with LLMs before seeing a medical professional, in a similar way that some patients use search engines before consulting a physician today. This poses the risk of fostering misguided self-diagnoses, with potentially harmful health consequences, especially in settings where access to health care can be financially challenging. Furthermore, patients might develop unrealistic expectations or demand unnecessary clinical resources, based on the information the LLM has provided. As a result, this could affect the dynamics of the patient-clinician relationship. On the other hand, LLMs possess the capacity to tailor medical information to the patient's level of comprehension, which may help enhance therapeutic education and adherence to medical advice.

Mitigating Hallucinations in Clinical Practice

Hallucinations can therefore represent a significant source of error if unaccounted for when using LLMs. A proactive and systematic approach is necessary to help interpret LLM output data and avoid succumbing to avoidable pitfalls that could cause harm to patients. This approach is summarized in Textbox 1.

Textbox 1. Important technical considerations before integrating large language models (LLMs) into clinical practice.

- What dataset was the LLM trained upon?
  - What specific considerations does this entail, with regard to bias?
  - Is the dataset up to date?
- What organization is behind the LLM and the dataset?
- Is the LLM specifically tailored for medical purposes?
- In testing rounds:
  - Is the information given by the LLM consistent with the existing knowledge on the subject?
  - Are the recommendations made by the LLM adequate compared to the accepted standard of care?
- How much variability exists within the LLM's responses? Can it be clinically significant for patient care?
- Enhanced capacity:
  - Does the model possess the capacity to integrate up-to-date information?
  - Does the model possess the capacity to search within reliable sources of information to better respond to the user's request? If the answer to one of the two above questions is yes, is this feature integrated within the LLM or is it operated by a third-party plug-in?
    - If a third-party plug-in is involved, what strengths and shortcomings does it entail?
  - Does the LLM possess the capacity to assess its answers' reliability?
  - Is the LLM capable of providing the links to its sources of information?
- Do best practice guidelines exist for the utilization of the LLM?
  - If so, are they specific to use in the health care sector?
  - If the guidelines are general use or specifically focused on another industry sector, what precautions must be applied before extrapolating their use to the health care sector?
- Has the model been tested in a rigorous fashion, and are the results of this evaluation subject to scientific publication?

XSL•FO
RenderX

First and foremost, understanding the model's origins, version, training database content, and strengths as well as drawbacks are essential prerequisites for an informed use of the LLM. With this information, the user should actively seek out what types of bias the model may contain and understand how it can affect the LLM's answers [2]. Furthermore, information on the training dataset should be sought out to understand how up-to-date the knowledge within it is, as well as if it is well equipped to answer medical inquiries. In this regard, a topic of emerging importance is the development of LLMs specifically trained for medical purposes. Although theoretically more performant than general LLMs, their relevance for clinical practice has not yet been evaluated.

Second, user input should be carefully crafted to create a high-quality request. The request should contain a detailed description of the clinical context; this requires carrying out a thorough history, clinical exam, and incorporating current as well as historical workup data. Clinical acumen thus remains essential in creating an adequate request. Therefore, although initially time-consuming, a higher-quality request can yield a more relevant answer.

Third, model accuracy and hallucination prevalence should be assessed before being put into practice, through iterative testing and evaluation. During testing rounds, LLM accuracy should be examined using standardized scenarios. Consistency, as well as variability in the answers, should be evaluated by regenerating the LLM's responses multiple times. Whether through formal, statistical evaluation or through getting a general sense of the model's characteristics, the clinician can evaluate the LLM's capabilities and shortcomings in this manner. Furthermore, the scientific adequacy of the LLM's responses should be assessed with regard to current standards of care and up-to-date guidelines. Even without knowing the database's knowledge cutoff, this method can help assess how up-to-date the data are, as well as understand how often hallucinations arise with regard to a specific subject.

Another useful tool that can help the clinician assess the reliability of the LLM's responses is plug-ins. These are usually third-party apps that can be programmed to serve a wide range of functions, including but not limited to, searching the internet, retrieving information from scientific databases, and substantiating responses with links to the sources of the presented information. LLMs may also possess certain of these capacities directly within the scope of their own functions. Although plug-ins may significantly enhance the reliability of the LLM's responses, by providing the ability for up-to-date referencing, they are not a guarantee that the response will be free from hallucinations. Therefore, plug-ins should be evaluated with the same amount of scrutiny as the LLM itself.

Finally, combining text, image, and video data in LLM training databases can lead to more accurate responses and may decrease the likelihood of hallucinations [20]. However, it is also important to remember that the multimodal model's performances still rely on the quality of their training dataset [21].

A crucial aspect in the deployment and judicious implementation of LLMs in a clinical setting lies in the establishment of a proactive error reporting program. In conjunction with the aforementioned recommendations, the implementation of such a program facilitates the identification and reporting of near-miss incidents. At the individual level, this allows the user to develop a personal appreciation of the model's shortcomings, as well as the topics subject to hallucinations. On an institutional level, it can help develop best practice guidelines by identifying frequent hallucination presentations and more general errors. If LLM solutions are delivered as an on-premises solution, it is conceivable that error reporting will help fine-tune local models.

Prospects

Research in AI will largely contribute to reducing hallucinations, be it through fine-tuning of the underlying model, prompt engineering techniques, development of specific medical LLMs, or other innovative approaches.

Given the increasing awareness of hallucinations, and understanding the risk they potentially pose to patient safety, ingenious mitigation strategies have recently developed. Measurement of semantic entropy [22], algorithmic approaches to address root causes of hallucinations [23], and more classical methods such as Retrieval-Augmented Generation [24] are some of the many techniques that have been proposed to identify and reduce hallucinations. Ideally, an automated combination of different strategies may help both accurately identify and reduce the occurrence of hallucinations. This would help ensure LLM response accuracy and reliability.

To remain up-to-date with these rapid and substantial developments in technology, many approaches will be required to ensure clinicians stay current and use these tools to the best of their capacity. Specific health care–related research will be required to evaluate the full extent of individual LLM capacities and performance. Furthermore, in the same way clinicians require continuing education in emerging and evolving health care topics, frequent training will be essential to use LLM-related tools adequately. Particular attention should be directed towards recognizing and mitigating the numerous pitfalls associated with their use. To this end, we have identified significant practical limitations of LLMs that may limit their rapid uptake into daily clinical practice (Textbox 2).

XSL•FO
RenderX

Textbox 2. Factors currently limiting the widespread implementation of large language models (LLMs) in clinical settings.

Model related

- Risk of bias related to:
  - The quality and quantity of the data used to train the model
  - How the model was trained and especially fine-tuned
  - How the LLM reacts to different methods of prompting
- Difficulty of complex models to explicit the reasoning behind their responses
  - Hallucination risk, without the capacity to inform the user on the final output's trustworthiness: information related to the source of the information not necessarily provided
  - No assessment of source reliability

Human related

- Risk of misuse related to:
  - Idealization of LLM capacities, and considering them to be completely foolproof, may lead to expert bias
  - Biased user input and uncritical approval of LLM output, aligning with the user's anticipated answer, can result in confirmation bias
  - Increased dependency on automated aids without critical thinking and reassessment, can lead to automation bias
- Absence of a strong legal framework defining the scope and regulatory environment of LLMs, as of date
- Lack of institutional governance defining the following:
  - Methods of informing patients and obtaining their consent for use of LLMs in their health care pathway
  - Integration of LLM use within a legal framework
  - Accountability in case of an error resulting from the use of the LLM
  - Methods of initial testing, implementation, and continuous improvement of the selected LLM
  - General operating conditions, namely:
    - Defining specific tasks for which the LLM should be used
    - Ensuring a human has the final word in the decision-making process, even though it is assisted by the LLM
  - Training procedures and certifications required for health care professionals to use the tool
  - Implementation format (on-premises vs outsourced)
  - Quality control processes
  - Traceability of feedback and changes in LLM implementation
  - Ethical framework

Economically related

- Cost of the initial investment
- Cost related to maintenance in an "on-premises" format, including:
  - Employee wages
  - Infrastructure costs
  - Energy costs
- Concerns regarding sustainability, given the high energy consumption of servers used to power and train current LLM models

As previously mentioned, future iterations of LLMs could also be specifically trained on medical datasets and fine-tuned by expert clinician input. This could be an effective method of reducing hallucinations and would allow the tailoring of LLMs to specific fields of medicine.

Conclusions

Due to its vast potential, LLM integration into routine clinical practice is no longer a question of if, but when. As technology advances, the integration of LLMs with other AI tools

XSL•FO
RenderX

possessing multimodal analysis capabilities (text, audio, and image) will follow suit. These advances offer significant opportunities in terms of patient care. However, robust legal frameworks will be necessary to guide their use on a national scale, and institutional governance is key to their implementation for everyday use. Indeed, an informed approach to using these tools, as well as significant efforts in terms of capacity-building, are primordial to avoid falling victim to their well-identified shortcomings.

Acknowledgments

No generative artificial intelligence (AI) tools were used in the creation of this manuscript.

Authors' Contributions

DR and FB contributed to writing–original draft and writing–review and editing.

Conflicts of Interest

None declared.

References

1. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. arXiv Preprint posted online on June 12, 2017 [FREE Full text] [doi: 10.48550/arXiv.1706.03762]
2. Li H, Moon JT, Purkayastha S, Celi LA, Trivedi H, Gichoya JW. Ethics of large language models in medicine and medical research. Lancet Digit Health 2023;5(6):e333-e335 [FREE Full text] [doi: 10.1016/S2589-7500(23)00083-3] [Medline: 37120418]
3. Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a radiology board-style examination: insights into current strengths and limitations. Radiology 2023 Jun;307(5):307(5)-e230582. [doi: 10.1148/radiol.230582] [Medline: 37191485]
4. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health 2023 Apr;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
5. Ayers JW, Poliak A, Dredze M, Leas EC, Zhu Z, Kelley JB, et al. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA Intern Med 2023 Jun 01;183(6):589-596 [FREE Full text] [doi: 10.1001/jamainternmed.2023.1838] [Medline: 37115527]
6. Kung TH, Cheatham M, Medenilla A, Sillos C, de Leon L, Elepaño C, et al. Performance of chatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023 Feb 09;2(2):e0000198 [FREE Full text] [doi: 10.1371/journal.pdig.0000198] [Medline: 36812645]
7. Dziri N, Milton S, Yu M, Zaiane O, Reddy S. On the origin of hallucinations in conversational models: is it the datasets or the models? arXiv:2204.07931 2022 Apr 17:1-14. [doi: 10.18653/v1/2022.naacl-main.387]
8. Ji Z, Lee N, Frieske R, Yu T, Su D, Xu Y, et al. Survey of hallucination in natural language generation. ACM Comput Surv 2023 Mar 03;55(12):1-38. [doi: 10.1145/3571730]
9. Azaria A, Mitchell T. The internal state of an LLM knows when its lying. 2023 Presented at: Findings of the Association for Computational Linguistics: EMNLP 2023; December 6-10, 2023; Singapore p. 967-976. [doi: 10.18653/v1/2023.findings-emnlp.68]
10. Yin Z, Sun Q, Guo Q, Wu J, Qiu X, Huang X. Do large language models know what they don't know? arXiv Preprint posted online on May 29, 2023 [FREE Full text] [doi: 10.48550/ARXIV.2305.18153]
11. Feldman P, Foulds JR, Pan S. Trapping LLM hallucinations using tagged context prompts. arXiv Preprint posted online on June 9, 2023 [FREE Full text] [doi: 10.48550/ARXIV.2306.06085]
12. Patel SB, Lam K. ChatGPT: the future of discharge summaries? Lancet Digit Health 2023 Mar;5(3):e107-e108 [FREE Full text] [doi: 10.1016/S2589-7500(23)00021-3] [Medline: 36754724]
13. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. Semin Ophthalmol 2023;38(5):503-507. [doi: 10.1080/08820538.2023.2209166] [Medline: 37133418]
14. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. N Engl J Med 2023 Mar 29;388(13):1233-1239. [doi: 10.1056/NEJMsr2214184] [Medline: 36988602]
15. Shea YF, Lee CMY, Ip WCT, Luk DWA, Wong SSW. Use of GPT-4 to analyze medical records of patients with extensive investigations and delayed diagnosis. JAMA Netw Open 2023;6(8):e2325000 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.25000] [Medline: 37578798]
16. Chen TC, Kaminski E, Koduri L, Singer A, Singer J, Couldwell M, et al. Chat GPT as a neuro-score calculator: analysis of a large language model's performance on various neurological exam grading scales. World Neurosurg 2023;179:e342-e347. [doi: 10.1016/j.wneu.2023.08.088] [Medline: 37634667]
17. McGowan A, Gui Y, Dobbs M, Shuster S, Cotter M, Selloni A, et al. ChatGPT and Bard exhibit spontaneous citation fabrication during psychiatry literature search. Psychiatry Res 2023 Aug;326:115334. [doi: 10.1016/j.psychres.2023.115334] [Medline: 37499282]

XSL•FO
RenderX

18. Wu RT, Dang RR. ChatGPT in head and neck scientific writing: a precautionary anecdote. Am J Otolaryngol 2023 Jul 06;44(6):103980. [doi: 10.1016/j.amjoto.2023.103980] [Medline: 37459740]
19. Medical practice variation: background paper. Australian Commission on Safety and Quality in HealthCare (ACSQHC). 2013. URL: https://www.safetyandquality.gov.au/sites/default/files/migrated/SAQ110_Medical_Practice_variation_V10_WEB.pdf [accessed 2024-04-01]
20. Chen J, Gui C, Ouyang R, Gao A, Chen S, Chen GH, et al. HuatuoGPT-vision, towards injecting medical visual knowledge into multimodal LLMs at scale. 2024 Presented at: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing; November 12-16, 2024; Miami, FL p. 7346-7370. [doi: 10.18653/v1/2024.emnlp-main.418]
21. Yin S, Fu C, Zhao S, Li K, Sun X, Xu T, et al. A survey on multimodal large language models. Natl Sci Rev 2024 Dec 23;11(12):nwae403 [FREE Full text] [doi: 10.1093/nsr/nwae403] [Medline: 39679213]
22. Farquhar S, Kossen J, Kuhn L, Gal Y. Detecting hallucinations in large language models using semantic entropy. Nature 2024 Jun 19;630(8017):625-630 [FREE Full text] [doi: 10.1038/s41586-024-07421-0] [Medline: 38898292]
23. Zhou Y, Cui C, Yoon J, Zhang L, Deng Z, Finn C, et al. Analyzing and mitigating object hallucination in large vision-language models. arXiv Preprint posted online on October 1, 2023 [FREE Full text] [doi: 10.48550/arXiv.2310.00754]
24. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. arXiv Preprint posted online on May 22, 2020 [FREE Full text] [doi: 10.48550/arXiv.2005.11401]

Abbreviations

AI: artificial intelligence
LLM: large language model

Edited by T de Azevedo Cardoso; submitted 24.04.24; peer-reviewed by J Aarts, J Luo, A Hidki; comments to author 10.09.24; revised version received 17.10.24; accepted 14.11.24; published 28.01.25.

Please cite as:
Roustan D, Bastardot F
The Clinicians' Guide to Large Language Models: A General Perspective With a Focus on Hallucinations
Interact J Med Res 2025;14:e59823
URL: https://www.i-jmr.org/2025/1/e59823
doi:10.2196/59823
PMID:

©Dimitri Roustan, François Bastardot. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.01.2025. 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

An Automated Clinical Laboratory Decision Support System for Test Utilization, Medical Necessity Verification, and Payment Processing

Safedin Beqaj[1], PhD; Rojeet Shrestha[2], PhD; Tim Hamill[3], MD

[1]Medical Database, Inc, Irvine, CA, United States

[2]Patient Choice Laboratories, Indianapolis, IN, United States

[3]Department of Laboratory Medicine, University of California, San Francisco, CA, United States

Corresponding Author:
Safedin Beqaj, PhD
Medical Database, Inc
1 Post, Suite 200
Irvine, CA, 92618
United States
Phone: 1 8477693701
Email: sajo@medicaldatabase.com

Abstract

Physicians could improve the efficiency of the health care system if a reliable resource were available to aid them in better understanding, selecting, and interpreting the diagnostic laboratory tests. It has been well established and widely recognized that (1) laboratory testing provides 70%-85% of the objective data that physicians use in the diagnosis and treatment of their patients; (2) orders for laboratory tests in the United States have increased, with an estimated volume of 4-5 billion tests per year; (3) there is a lack of user-friendly tools to guide physicians in their test selection and ordering; and (4) laboratory test overutilization and underutilization continue to represent a pervasive source of inefficiency in the health care system. These inappropriate test orders not only lead to slower or incorrect diagnoses for patients but also add a significant financial burden. In addition, many ordered tests are not reimbursed by Medicare because they are inappropriate for the medical condition or were ordered with the incorrect International Statistical Classification of Diseases and Related Health Problems, Tenth Revision diagnostic code, not meeting the medical necessity. Therefore, current clinical laboratory test ordering procedures experience a quality gap. Often, providers do not have access to an appropriate tool that uses evidence-based guidelines or algorithms to ensure that tests are not duplicated, overused, or underused. This viewpoint lays out the potential use of an automated laboratory clinical decision support system that helps providers order the right test for the right disease and documents the right reason or medical necessity to pay for the testing.

(Interact J Med Res 2025;14:e46007) doi:10.2196/46007

KEYWORDS

clinical decision system; CDSS; laboratory decision system; laboratory testing; test utilization; test ordering; lab test; laboratory; testing; payment; decision-making; user; utilization; processing; decision

Introduction

Laboratory testing plays a key role in clinical decision-making and physician orders for laboratory tests are increasing [1,2]. It is estimated that at least 20% of the 4-5 billion lab orders submitted annually in the United States are inappropriate. Studies have shown that overutilization and underutilization of laboratory tests occur 20.6% and 44.8% of the time, respectively [3]. This inappropriate testing not only leads to incorrect or delayed diagnoses but also significantly adds a financial burden on the health care system. This situation is expected to worsen as the available lab tests menu grows, especially in the areas of molecular diagnostics and genetic testing. Due to a lack of physician test information, education, and insurance coverage questions, ordering less effective and sometimes obsolete tests over newer tests that are more sensitive and specific remains a major problem [4]. This inappropriate testing not only led to incorrect or delayed diagnoses but also significantly added financial burden. The situation is expected to get worse as the number of lab tests is growing, especially in molecular diagnostics and genetic testing. The introduction of an automated

XSL•FO
RenderX

clinical decision support system (CDSS) that guides physicians to order the most appropriate test(s) for their patients while simultaneously providing both medical necessity requirements and applicable diagnostic codes will be a vital tool to improve test ordering and reimbursement efficiency.

Medicare and commercial health-care plans all require that ordered tests are accompanied by appropriate *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (*ICD-10*) diagnostic codes that meet medical necessity rules. These requirements make it a complex process for providers to decide which tests to order, provide diagnostic information, and obtain previous authorization if required, so that test bills and payments are efficient and timely [5,6]. Given the rapidly growing demand for tests, especially molecular and genetic testing, the lack of a reliable laboratory CDSS will compound this already complex process.

When physicians fail to select and order the most appropriate test(s) based on the patient's health condition and further fail to provide the proper diagnostic codes to support medical necessity, laboratory billing will most certainly fail. The patient may then be held responsible for the laboratory charges and the laboratory will be caught in the middle of disagreements between the insurance company, the treating physician, and the patient in determining the responsible party for the laboratory charges. Ideally, every test ordered and procedure performed by the lab should be paid or reimbursed by health insurance. However, many ordered tests are not reimbursed, primarily due to a lack of medical necessity. This issue arises from ordering the wrong test that does not meet medical necessity criteria or failing to provide the correct diagnostic code for the disease or health condition. Therefore, the current clinical laboratory test-ordering procedures suffer from a quality gap and require an automated system to address this issue.

This viewpoint discusses the use of an automated laboratory CDSS that helps providers order the right test for the right disease and documents the right reason or medical necessity to pay for the testing.

Inappropriate Test Ordering

Inappropriate testing encompasses both overutilization and underutilization, both of which can affect quality patient care and health care expenditures. Overutilization includes tests that are ordered but not indicated, tests that are ordered at the incorrect time in the clinical course, or tests that are ordered too frequently. Conversely, underutilization refers to tests that are indicated but not ordered, or those that are not ordered at the appropriate time to positively impact patient care [2,5]. Both can have an adverse impact on the quality of patient care and health care costs because of downstream consequences such as additional diagnostic testing, repeat testing, imaging, prescriptions, surgeries, or prolonged hospital stays. It is estimated that more than three hundred million patients visit the laboratory annually and that at least 23 million of these patients are affected by inappropriate test ordering and test interpretation. The reports on the Commonwealth Fund Survey of Public Views of the US Health Care System (2012) found that more than 23% of laboratory tests ordered by physicians were duplicated or repeated, which increases the cost of care by further delaying or confusing the patient's diagnosis and care. It is also reported that overutilization and underutilization of laboratory tests occurred 20.6% and 44.8% of the time, respectively [2].

In 2011, a survey conducted by the Centers for Disease Control and Prevention among primary care physicians found that 14.7% had uncertainty in selecting and ordering the correct test and 8.3% had difficulty interpreting tests [7]. Physicians' lack of access to information about the availability of specific tests, the limitations of tests, and insurance coverage might play a major role in the underutilization of tests, whereas medical malpractice, selecting obsolete testing, and use of certain aspects of computerized provider order entry are documented factors leading to overutilization [3,4].

Economic Impact of Test Misutilization

Expenditures for health care in the US were approximately $4.1 trillion in 2020, which is an increase of 9.7% from 2019, and it accounts for 19.7% of the total gross domestic product. If the trends continue, health care costs are projected to increase to $6.2 trillion by 2028 [7]. Hospital and clinical service expenditure also showed rapid growth in 2020 and accounted for approximately $2.08 trillion of the total cost of health care. Although laboratory testing accounts for only a fraction of health care expenditures, 94% of objective and structured data in the electronic medical record (EMR) are obtained from a clinical laboratory [8]. Moreover, it is estimated that 60%-70% of all clinical decisions are based on the results of laboratory testing [9,10]. Considering 60% as the rate of influence on the clinical decision, it can be estimated that $1.2 trillion of health care spending is influenced by laboratory testing. Therefore, inappropriate testing not only leads to incorrect or delayed diagnoses but also significantly adds financial burden. 4-5 billion tests are performed annually in the United States. Unfortunately, it is estimated that at least 20% of the lab orders submitted are inappropriate [2]. The situation is expected to get worse as the number of esoteric lab tests is growing, especially in the areas of molecular diagnostics and genetic testing.

Ideally, every test performed by the laboratory should be reimbursed. However, many billed tests are not reimbursed due to a lack of documentation ensuring medical necessity. In many cases, the denial of the reimbursement is due to the submission of improper diagnostic code(s) for the disease or health problem being tested for. Therefore, current clinical laboratory test ordering procedures experience an information gap and there is an urgent need for an automated system to improve test utilization for economic sustainability in health care.

Need for Clinical Decision Support System in Clinical Laboratory

Selecting and obtaining authorization for appropriate medical tests is an ongoing and growing challenge in many specialties, including radiology, cardiology, pulmonology, and pharmacology. With typical radiology and diagnostic imaging costs higher than those for laboratory testing, the US government

XSL•FO
RenderX

has prioritized approval of a reimbursement reward system for insurance providers that use a CDSS to improve imaging utilization. For example, there is a 2015 "Advanced Imaging Bill" which mandates that government-approved imaging services will only be reimbursed if the insurance claim confirms that appropriate-use criteria were consulted or a CDSS was used. The bill also recommends the use of CDSS for other diagnostic test ordering, if available [5,11,12].

CDSS is currently also available in cardiology, medication management, oncology, and urology. These broad and growing applications along with expansive and expensive specialized lab testing strongly indicate that there is a substantial need for an expert laboratory CDSS to aid health care providers, care managers, and payers in selecting, ordering, and approving laboratory tests and reducing inappropriate testing.

Currently, there are some partially developed and semimanual lab CDSSs that help physicians order laboratory tests; however, these approaches are provider-driven and require inconvenient interactive user questions to access the information needed [10]. Unlike radiology CDSSs, these systems do not provide any scoring system for tests based on medical evidence, clinical relevancy, or medical necessity. Incorporating a scoring system based on test indications and providing information on supportive diagnostic codes can help automate the laboratory test ordering process and has positive impacts on test utilization, medical necessity documentation, claim verification, and payment processing. These developments strongly indicate that there is also a substantial need for a laboratory CDSS to help health care providers in selecting and ordering the appropriate laboratory tests, reduce inappropriate testing, aid providers in easier and more automated payment processing, and finally get better and on-time health care to patients [4,5,13,14].

Solution to This Problem

A potential solution is to develop a laboratory CDSS that will aid providers in selecting and ordering the right diagnostic tests with which to manage patient health care. The CDSS will help laboratories process the order, process the sample, and report accurate results on-time delivery to the ordering provider. The CDSS will provide information regarding the appropriate diagnostic *ICD-10* code(s) to meet the medical necessity. The CDSS will also provide a medical evidence-based scoring system based on clinical utility. A CDSS that provides the testing indication(s) to complement the provider's notes and is electronically interfaced with EMRs, electronic hospital records, and billing systems to automate processes like medication management and radiology CDSS is desirable [4,5,14].

Laboratory Decision System (LDS), developed by Medical Database Inc, is one of the available automated CDSS. LDS is an algorithm-based test selection and ordering database for physicians, health care providers, insurance and managed care companies, and billing services. LDS is expertly developed to help system users understand, select, order, and use laboratory tests for disease diagnosis and management using evidence-based guidelines and industry best practices. The system uses our proprietary scoring system developed by our editorial board (60 pathologists and PhDs), designed to rank testing recommendations based on disease, clinical relevance, medical necessity, and testing indication. Each time an order is placed via the LDS platform, it automatically includes the appropriate diagnostic *ICD-10* code and Local Coverage Determinations (LCD) or National Coverage Determinations (NCD) to meet medical necessity for reimbursement. Included in the robust database are all commercially available tests (over 2300 diagnostic tests), including genetic and proprietary tests [5].

The LDS rates and scores potential tests for any given disease and assigns an easily interpretable numerical and color-coded score based on clinical relevance, medical necessity, and testing indication. Tests with scores of 5 or above (10 being the highest score) meet medical necessity, while those with scores of 4 or less do not. LDS also follows Medicare's medical necessity guidelines by using testing indications such as "initial testing indication" to allow providers to better characterize the patient's disorder based on initial test results before ordering overly complex and expensive tests. Appropriate tests use indication labels, for example, diagnostic, disease management, monitoring, and alternative tests, categorizing each test with the right indication or reason for testing to avoid using providers' charts and notes that make it difficult to automate the system [4,5]. When assessing the effectiveness of LDS in improving test utilization and reimbursement with 96,170 laboratory requests comprising 374,423 test orders from a reference laboratory, 44,671 tests were accompanied by *ICD-10* that are described by Medicare as "never covered" because of the lack of a system to check or support the medical necessity of each order. A total of 160,449 tests were subject to a Medicare policy review from which 112,400 tests met coverage criteria, and 48,049 tests did not. These orders were then reevaluated using LDS. Of the original test order sample, 91.5% had an associated LDS score. Of these scored tests, 47.8% met coverage and 43.73% failed to meet coverage, according to the LDS Ranking System. Importantly, LDS provided recommendations for alternative diagnostic *ICD-10* codes or tests which could have aided physicians in choosing a more appropriate test or submitting a different *ICD-10* diagnostic code to meet medical necessity. Around 96.4% have an alternative *ICD-10* code or a test score above 5, meeting medical necessity. The LDS system recommended 80.5% which would meet Medicare policies, demonstrating that the LDS system would correct inappropriate orders if used as a testing utilization management system [5]. However, more systemic testing of the platform is needed to evaluate the effectiveness of test utilization, medical necessity verification, and payment processing.

Since the LDS platform has been built to interface with EMRs, electronic hospital records, and laboratory information systems (LIS), the content can be accessed directly through these systems, allowing orders to be sent directly to laboratories for testing coordination. Accordingly, when using the LDS platform, every test ordered will automatically include a medical necessity score, the correct testing indication, and the appropriate *ICD-10* and CPT codes, all of which also support adjudication for bill payment. An outline of the use of computerized provider order entry of LDS is elucidated in Figure 1. In addition, each test

XSL•FO
RenderX

ordered through the LDS platform will provide testing indications to support the purpose of the testing, thereby reducing manual submission of "medical necessity" review data, including reasons for test ordering (scripts, notes, charts, etc) and adding system automation with lower costs, faster throughput, and higher performance [4,15].

Figure 1. Test ordering using LDS automates the selection of appropriate tests based on clinical relevance and integrates the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision and LCD codes to facilitate reimbursement. LCD: Local Coverage Determinations; LOINC: Logical Observation Identifiers Names and Codes; NCD: National Coverage Determinations.



XSL•FO
RenderX

Conclusion

In conclusion, there is a clear and immediate need for an LDS system similar to that which is used in radiology and medication management, which can aid providers in selecting the right test for each disease or condition while assigning the correct *ICD-10* code, right Local Coverage Determinations and National Coverage Determinations to meet the medical necessity and right testing indication covering the reason and use of ordered test(s). The available LDS system developed by Medical Database, described in this viewpoint study may assist providers in making appropriate utilization decisions while also supporting laboratories in reimbursement and streamlining claim verification for payers, all of which combined will potentially make the laboratory industry and overall health care more efficient and cost-effective.

Conflicts of Interest

SB is the founder and CEO of Medical Database, Inc and developed the Laboratory Decision System.

References

1. Jackson BR. Managing laboratory test use: principles and tools. Clin Lab Med 2007;27(4):733-748. [doi: 10.1016/j.cll.2007.07.009] [Medline: 17950895]
2. Zhi M, Ding EL, Theisen-Toupal J, Whelan J, Arnaout R. The landscape of inappropriate laboratory testing: a 15-year meta-analysis. PLoS One 2013;8(11):e78962 [FREE Full text] [doi: 10.1371/journal.pone.0078962] [Medline: 24260139]
3. Yeh DD. A clinician's perspective on laboratory utilization management. Clin Chim Acta 2014;427:145-150. [doi: 10.1016/j.cca.2013.09.023] [Medline: 24084504]
4. Hickner J, Thompson PJ, Wilkinson T, Epner P, Sheehan M, Pollock AM, et al. Primary care physicians' challenges in ordering clinical laboratory tests and interpreting results. J Am Board Fam Med 2014;27(2):268-274 [FREE Full text] [doi: 10.3122/jabfm.2014.02.130104] [Medline: 24610189]
5. Leblow L, Hamill T, Beqaj SH. Use of laboratory decision system as a test utilization management tool in clinical settings, current and future perspectives. J Clin Lab Med 2019;4(1). [doi: 10.16966/2572-9578.128]
6. Sireci AN, Patel JL, Joseph L, Hiemenz MC, Rosca OC, Caughron SK, et al. Molecular pathology economics 101: an overview of molecular diagnostics coding, coverage, and reimbursement: a report of the association for molecular pathology. J Mol Diagn 2020;22(8):975-993 [FREE Full text] [doi: 10.1016/j.jmoldx.2020.05.008] [Medline: 32504675]
7. NHE fact sheet. Centers for Medicare & Medicaid Services. URL: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet [accessed 2024-12-18]
8. Forsman R. The electronic medical record: implications for the laboratory. Clin Leadersh Manag Rev 2000;14(6):292-295. [Medline: 11210217]
9. Rohr UP, Binder C, Dieterle T, Giusti F, Messina CGM, Toerien E, et al. The value of in vitro diagnostic testing in medical practice: a status report. PLoS One 2016;11(3):e0149856 [FREE Full text] [doi: 10.1371/journal.pone.0149856] [Medline: 26942417]
10. Hallworth MJ. The '70% claim': what is the evidence base? Ann Clin Biochem 2011;48(Pt 6):487-488. [doi: 10.1258/acb.2011.011177] [Medline: 22045648]
11. Bellodi E, Vagnoni E, Bonvento B, Lamma E. Economic and organizational impact of a clinical decision support system on laboratory test ordering. BMC Med Inform Decis Mak 2017;17(1):179 [FREE Full text] [doi: 10.1186/s12911-017-0574-6] [Medline: 29273037]
12. Fornell D. CMS to require appropriate use criteria documentation for medical imaging orders. Diagnostic and Interventional Cardiology. URL: https://www.dicardiology.com/article/cms-require-appropriate-use-criteria-documentation-medical-imaging-orders [accessed 2017-04-18]
13. McMorris Rodgers C. https://www.congress.gov/bill/114th-congress/house-bill/2550/text. Congress.gov. 2015 May 15. URL: https://www.congress.gov/bill/114th-congress/house-bill/2550/text [accessed 2025-02-06]
14. Baron JM, Dighe AS. The role of informatics and decision support in utilization management. Clin Chim Acta 2014;427:196-201. [doi: 10.1016/j.cca.2013.09.027] [Medline: 24084507]
15. Delvaux N, Van Thienen K, Heselmans A, de Velde SV, Ramaekers D, Aertgeerts B. The efects of computerized clinical decision support systems on laboratory test ordering: a systematic review. Arch Pathol Lab Med 2017;141(4):585-595. [doi: 10.5858/arpa.2016-0115-RA] [Medline: 28353386]

Abbreviations

CDSS: clinical decision support system
EMR: electronic medical record
ICD-10: International Statistical Classification of Diseases and Related Health Problems, Tenth Revision
LDS: Laboratory Decision System

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 25.01.23; peer-reviewed by C Schäfer; comments to author 29.01.24; revised version received 13.10.24; accepted 14.01.25; published 11.02.25.

Please cite as:
Beqaj S, Shrestha R, Hamill T
An Automated Clinical Laboratory Decision Support System for Test Utilization, Medical Necessity Verification, and Payment Processing
Interact J Med Res 2025;14:e46007
URL: https://www.i-jmr.org/2025/1/e46007
doi:10.2196/46007
PMID:39808833

©Safedin Beqaj, Rojeet Shrestha, Tim Hamill. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 11.02.2025. 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

Use of Clinical Public Databases in Hidradenitis Suppurativa Research

Xu Liu[1,2], MS; Linghong Guo[1,2*], PhD; Xian Jiang[1,2*], PhD

[1]Department of Dermatology, West China Hospital, Sichuan University, Chengdu, China

[2]Laboratory of Dermatology, Clinical Institute of Inflammation and Immunology, Frontiers Science Center for Disease-Related Molecular Network, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Xian Jiang, PhD
Department of Dermatology
West China Hospital
Sichuan University
#37 Guoxue Alley
Wuhou District
Chengdu, 610041
China
Phone: 86 028 85423315
Fax: 86 028 85422560
Email: jiangxian@scu.edu.cn

Abstract

In this viewpoint, we argue that recent studies using clinical public databases have revolutionized our understanding of hidradenitis suppurativa (HS), a chronic inflammatory skin condition with significant impacts on patients' quality of life. Our key messages are as follows: (1) these databases enable large-scale studies integrating genetic, epidemiological, and clinical data, providing crucial insights into HS's genetic predispositions, comorbidities, and treatment outcomes; (2) findings highlight a strong genetic component, with mutations in the γ-secretase complex playing a key role in HS pathogenesis and shaping targeted therapies; (3) studies also reveal elevated risks for comorbidities like obesity, diabetes, cardiovascular disease, and systemic inflammation in patients with HS, with diet-driven inflammatory pathways potentially exacerbating disease severity; (4) while these databases offer unprecedented research opportunities, limitations such as data representativeness and quality must be considered; (5) nonetheless, their benefits outweigh potential drawbacks, allowing the identification of rare comorbidities, disease progression patterns, and personalized treatment strategies; and (6) increased funding for HS research is crucial to harness these databases' full potential, develop targeted therapies, and ultimately improve patient outcomes. As HS's impact is disproportionate to current research investments, we believe advocating for more resources and addressing database limitations will be key to advancing HS understanding and care.

(Interact J Med Res 2025;14:e70282) doi:10.2196/70282

KEYWORDS

hidradenitis suppurativa; clinical public databases; disease progression; patient data; HS

Hidradenitis suppurativa (HS) is a debilitating chronic inflammatory skin condition that significantly impacts patients' quality of life. Despite its profound effects, the pathogenesis of HS remains poorly understood [1]. In this viewpoint, we aim to highlight how recent advancements in clinical public databases have provided researchers with a powerful tool to explore the genetic, epidemiological, and clinical aspects of HS (Table 1). We argue that these databases have enabled large-scale studies integrating diverse patient data, yielding crucial insights into the disease's genetic predispositions, comorbidities, and treatment outcomes [2].

XSL•FO
RenderX

Table 1. Recent studies on hidradenitis suppurativa (HS) using clinical public databases.

Title	Year	Journal	Conclusion	Databases
Genetic Susceptibility to Hidradenitis Suppurativa and Predisposition to Cardiometabolic Disease [3]	2024	*JAMA Dermatology*	Explored the genetic susceptibility of HS to cardiovascular and metabolic diseases	UK Biobank
Association Between Hidradenitis Suppurativa and Gout: A Propensity Score-Matched Cohort Study [4]	2024	*Dermatology*	Found an association between HS and gout	TriNetX Research Network
A History of Asthma Is Associated With Susceptibility to Hidradenitis Suppurativa: A Population-Based Longitudinal Study [5]	2023	*Archives of Dermatological Research*	Found a correlation between a history of asthma and susceptibility to HS	Clalit Health Services
Hidradenitis Suppurativa and the Risk of Myocardial Infarction, Cerebrovascular Accident, and Peripheral Vascular Disease: A Population-Based Study [6]	2023	*Archives of Dermatological Research*	Investigated the risk of myocardial infarction, stroke, and peripheral vascular disease in patients with HS	Clalit Health Services
Hidradenitis Suppurativa and Rheumatoid Arthritis: Evaluating The Bidirectional Association [7]	2021	*Immunologic Research*	Evaluated the bidirectional association between HS and rheumatoid arthritis	Clalit Health Services
Association of Birth Weight, Childhood Body Mass Index, and Height With Risk of Hidradenitis Suppurativa [8]	2020	*JAMA Dermatology*	Studied the relationship between birth weight, childhood BMI, and height with the risk of HS	Danish National Patient Register
Global Hidradenitis Suppurativa COVID-19 Registry: A Registry to Inform Data-Driven Management Practices [9]	2020	*British Journal of Dermatology*	Explored data-driven management practices through the Global Hidradenitis Suppurativa COVID-19 Registry	Global Hidradenitis Suppurativa COVID-19 Registry
Comparing Cutaneous Research Funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases With 2010 Global Burden of Disease Results [10]	2014	*PLoS One*	Compared skin research funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases with the 2010 Global Burden of Disease results	National Institute of Arthritis and Musculoskeletal and Skin Diseases Database and the Global Burden of Disease

One of our key messages is that studies using clinical public databases have revealed a strong genetic component in HS. By analyzing data from diverse populations, researchers have identified genetic factors, such as mutations in the γ-secretase complex, that predispose certain individuals to develop HS [3,11]. We believe these genetic insights are not only expanding our understanding of HS pathogenesis but also shaping the development of targeted therapies [11]. Furthermore, we emphasize that clinical databases have shed light on the complex relationship between HS and various comorbidities. Studies have shown that patients with HS have a significantly higher risk of developing obesity, diabetes, cardiovascular disease, and even gout [4,6,12,13]. Notably, diet-driven inflammatory pathways, particularly those involving interleukin-17 and interleukin-1β, have been implicated in exacerbating HS severity [12]. The interplay between obesity, immune checkpoint inhibitors, and psoriasiform eruptions in HS further highlights the role of shared inflammatory pathways in disease exacerbation [13]. These findings underscore our view that a holistic approach to managing HS is needed, addressing both the dermatological and broader health risks faced by patients.

While we believe the insights gained from clinical public databases are invaluable, it is essential to consider potential limitations and counterarguments. One concern is the representativeness of the data within these databases. We argue that ensuring diverse and inclusive patient populations is crucial to avoid bias and generate findings applicable to the broader HS community [7]. Additionally, the quality and consistency of data entry across different health care settings may vary, potentially impacting the reliability of the results [5]. Furthermore, biases may exist in terms of the ethnic and geographic diversity of the data, which may skew the results for certain populations. We believe expanding the scope of these databases to include more diverse patient groups will be essential to ensure that the findings are broadly applicable and not limited by sample biases.

Despite these challenges, it is our view that the benefits of using clinical public databases in HS research far outweigh the limitations. These databases provide access to large, diverse patient populations and enable the identification of rare comorbidities and treatment outcomes that may not be apparent in smaller clinical settings. Moreover, longitudinal data captured in these databases allow researchers to study disease progression and identify early diagnostic markers, paving the way for timely interventions and personalized treatment plans [8]. We argue that researchers should focus on making improvements to database infrastructure, such as ensuring data consistency and addressing gaps in representation, to maximize their utility for future studies.

As we continue to harness the power of clinical public databases, we believe it is crucial to prioritize HS research and allocate adequate funding. The Global Burden of Disease study has

XSL•FO
RenderX

highlighted the significant impact of HS on patients' lives, yet research funding for HS remains disproportionately low compared to other skin diseases [9,10]. We argue that increased investment in HS research will accelerate the development of targeted therapies and improve patient outcomes. Public health initiatives aimed at increasing awareness of HS and its comorbidities, along with funding to support further research, will be vital to driving innovation and improving patient care.

In conclusion, it is our view that clinical public databases have revolutionized HS research by providing unprecedented access to large-scale, diverse patient data. The insights gained from these databases have deepened our understanding of the genetic susceptibility, comorbidities, and treatment outcomes associated with HS. As we move forward, we believe it is essential to address the limitations of these databases, ensure inclusive patient representation, and advocate for increased funding for HS research. By harnessing the full potential of clinical public databases, we can unlock new avenues for personalized medicine and ultimately improve the lives of individuals affected by this challenging condition.

Acknowledgments

This study was supported by the 1·3·5 Project for Disciplines of Excellence, West China Hospital, Sichuan University (ZYAI24028).

Data Availability

No new data were generated or analyzed in support of this viewpoint.

Authors' Contributions

LG (linhom.guo@foxmail.com) and XJ are cocorresponding authors.

Conflicts of Interest

None declared.

References

1. Jemec GB. Clinical practice. hidradenitis suppurativa. N Engl J Med 2012 Jan 12;366(2):158-164. [doi: 10.1056/NEJMcp1014163] [Medline: 22236226]
2. Goldburg SR, Strober BE, Payette MJ. Hidradenitis suppurativa: epidemiology, clinical presentation, and pathogenesis. J Am Acad Dermatol 2020 May;82(5):1045-1058. [doi: 10.1016/j.jaad.2019.08.090] [Medline: 31604104]
3. Nielsen VW, Bundgaard Vad O, Holgersen N, Paludan-Müller C, Meseguer Monfort L, Beyer AF, et al. Genetic susceptibility to hidradenitis suppurativa and predisposition to cardiometabolic disease. JAMA Dermatol 2025 Jan 01;161(1):22-30. [doi: 10.1001/jamadermatol.2024.3779] [Medline: 39382891]
4. Chang HC, Chiu TM, Tsai RY, Li CP, Wu YL, Chen SJ, et al. Association between hidradenitis suppurativa and gout: a propensity score-matched cohort study. Dermatology 2025 Oct 28;241(1):19-26. [doi: 10.1159/000541969] [Medline: 39467521]
5. Kridin K, Shihade W, Weinstein O, Zoller L, Onn E, Cohen A, et al. A history of asthma is associated with susceptibility to hidradenitis suppurativa: a population-based longitudinal study. Arch Dermatol Res 2023 Dec 29;315(10):2845-2851 [FREE Full text] [doi: 10.1007/s00403-023-02693-4] [Medline: 37642698]
6. Kridin K, Valido K, Cohen JM, Cohen AD. Hidradenitis suppurativa and the risk of myocardial infarction, cerebrovascular accident, and peripheral vascular disease: a population-based study. Arch Dermatol Res 2023 Apr 23;315(3):429-435. [doi: 10.1007/s00403-022-02369-5] [Medline: 35871185]
7. Kridin K, Shavit E, Damiani G, Cohen AD. Hidradenitis suppurativa and rheumatoid arthritis: evaluating the bidirectional association. Immunol Res 2021 Dec 19;69(6):533-540 [FREE Full text] [doi: 10.1007/s12026-021-09221-4] [Medline: 34409549]
8. Jørgensen AHR, Aarestrup J, Baker JL, Thomsen SF. Association of birth weight, childhood body mass index, and height with risk of hidradenitis suppurativa. JAMA Dermatol 2020 Jul 01;156(7):746-753 [FREE Full text] [doi: 10.1001/jamadermatol.2020.1047] [Medline: 32347905]
9. Naik HB, Alhusayen R, Frew J, Guilbault S, Ingram JR, Lowes MA, et al. Global Hidradenitis Suppurativa COVID-19 registry: a registry to inform data-driven management practices. Br J Dermatol 2020 Oct 02;183(4):780-781 [FREE Full text] [doi: 10.1111/bjd.19345] [Medline: 32597495]
10. Karimkhani C, Boyers LN, Margolis DJ, Naghavi M, Hay RJ, Williams HC, et al. Comparing cutaneous research funded by the National Institute of Arthritis and Musculoskeletal and Skin Diseases with 2010 Global Burden of Disease results. PLoS One 2014 Jul 8;9(7):e102122 [FREE Full text] [doi: 10.1371/journal.pone.0102122] [Medline: 25003335]
11. Chu Y, Yu S. Hidradenitis suppurativa: an understanding of genetic factors and treatment. Biomedicines 2024 Feb 01;12(2):338 [FREE Full text] [doi: 10.3390/biomedicines12020338] [Medline: 38397941]

XSL•FO
RenderX

12. Yu S, Wu X, Zhou Y, Sheng L, Jena PK, Han D, et al. A Western diet, but not a high-fat and low-sugar diet, predisposes mice to enhanced susceptibility to imiquimod-induced psoriasiform dermatitis. J Invest Dermatol 2019 Jun;139(6):1404-1407 [FREE Full text] [doi: 10.1016/j.jid.2018.12.002] [Medline: 30571973]
13. Yu S, Wu X, Shi Z, Huynh M, Jena PK, Sheng L, et al. Diet-induced obesity exacerbates imiquimod-mediated psoriasiform dermatitis in anti-PD-1 antibody-treated mice: implications for patients being treated with checkpoint inhibitors for cancer. J Dermatol Sci 2020 Mar;97(3):194-200 [FREE Full text] [doi: 10.1016/j.jdermsci.2020.01.011] [Medline: 32044178]

Abbreviations

HS: hidradenitis suppurativa

Edited by T de Azevedo Cardoso; submitted 18.12.24; peer-reviewed by X Li, S Yu; comments to author 17.01.25; revised version received 19.01.25; accepted 06.02.25; published 18.02.25.

Please cite as:
Liu X, Guo L, Jiang X
Use of Clinical Public Databases in Hidradenitis Suppurativa Research
Interact J Med Res 2025;14:e70282
URL: https://www.i-jmr.org/2025/1/e70282
doi:10.2196/70282
PMID:

©Xu Liu, Linghong Guo, Xian Jiang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.02.2025. 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

Effective Recruitment or Bot Attack? The Challenge of Internet-Based Research Surveys and Recommendations to Reduce Risk and Improve Robustness

Liesje Donkin[1], BSc, MSc(Hons), PGDipHlthPsych, PGDipArts(Dist), PGDipClinPsych, PhD; Nathan Henry[2], BTech, MSc; Amy Kercher[1], BPsych(Hons), GDipClinPsychSci, PhD; Mangor Pedersen[1], BA, MSc, PhD; Holly Wilson[3], BSocSci(Hons), MSc; Amy Hai Yan Chan[3], BPharm(Hons), PhD

[1]Department of Psychology and Neuroscience, Auckland University of Technology, Auckland, New Zealand

[2]School of Psychological Sciences, University of Western Australia, Perth, Australia

[3]School of Pharmacy, University of Auckland, Auckland, New Zealand

Corresponding Author:
Liesje Donkin, BSc, MSc(Hons), PGDipHlthPsych, PGDipArts(Dist), PGDipClinPsych, PhD
Department of Psychology and Neuroscience
Auckland University of Technology
Private Bag 92006
Auckland, 1142
New Zealand
Phone: 64 21847886
Email: liesje.donkin@aut.ac.nz

Abstract

Internet-based research has exploded in popularity in recent years, enabling researchers to offer both investigations and interventions to broader participant populations than ever before. However, challenges associated with internet-based research have also increased—notably, difficulties verifying participant data and deliberate data manipulation by bot and spam responses. This study presents a viewpoint based on 2 case studies where internet-based research was affected by bot and spam attacks. We aim to share the learnings from these experiences with recommendations for future research practice that may reduce the likelihood or impact of future attacks. The screening and verification processes used are presented and discussed, including the limitations of these. Based on our experience, security and screening within internet-based research platforms are partly effective, but no solution is available to protect researchers completely against bot attacks. Implications for future research and advice for health researchers are discussed.

(Interact J Med Res 2025;14:e60548) doi:10.2196/60548

KEYWORDS

internet-based research; research methodology; surveys; data integrity; bot attacks; technology; data manipulation; spam; false; falsification; fraudulent; fraud; bots; research methods; data collection; verify; verification; participants

Introduction

Overview

The internet and digital technologies have irretrievably changed the conduct of research. Whether it be participant engagement, digital intervention delivery, data collection, or distribution of findings, technology has allowed researchers to engage in wider-reaching recruitment. Technology has also enabled people to access and participate in research that may not have been previously possible. In particular, health researchers have embraced internet-based recruitment to invite participation from varied population groups easily and to collect data from large samples.

There has been a significant increase in sample sizes in the last decades, intending to increase statistical power, replicability, and generalizability of research findings [1]. The adoption of open science practices and the emergence of global collaborative efforts, such as the Human Connectome Project [2] and the UK Biobank study [3], including tens of thousands of participants, has facilitated the pooling of resources and increased the overall sample sizes in research studies. However, with the increase in public accessibility of internet-based studies comes a potential increase in poor quality or false participants, ranging from

XSL•FO
RenderX

careless respondents who respond with insufficient effort [4] to malicious responses such as automated attacks in the form of computer-programmed bots [5,6].

Bots can be fully automated to target internet-based research via automated malware or malicious algorithms [7], or they may be "hybrid" where there is an element of human control. An example of a hybrid bot is when the human may complete the initial survey screening questionnaires and then allow an automated bot or algorithm to complete the remainder of the survey. Bot attacks have existed since the early 2000s and are becoming increasingly common. This increase may be linked to the rise of internet-based paid research panels and crowdsource platforms such as Amazon Mechanical Turk, which bots can leverage. Whether the bots are automated malware or human respondents who are completing surveys for financial gain without meeting study inclusion criteria [8], it is irrefutable that caution must be exercised when conducting an internet-based or remote study [9,10].

Bot attacks must be identified as they impact the integrity of datasets [5,11,12], creating a situation where ineffective interventions may appear effective, and conversely, effective interventions may seem ineffective. Similarly, one of the significant risks of undetected bot attacks is the potential to misrepresent populations [5,13] and influence decision-making based on erroneous data. Misrepresentation is particularly problematic for vulnerable [14] and at-risk populations such as Indigenous communities, who are often poorly serviced by many existing interventions [15], or populations where new effective interventions are desperately needed [13,16]. Thus, misrepresentations of bots as members of this population could have dire consequences.

This paper discusses the viewpoints of strategies researchers can use to help reduce the impact of bots on research. Like several recent papers [5,13,17-21], the researchers were impacted by bot attacks on 2 independent projects hosted on different platforms at 2 institutions. We present case studies (CS1 and CS2) of the 2 research projects (see Multimedia Appendix 1) that were affected by bot attacks and demonstrate the impact these would have had on the demographics of the final dataset. We share the steps the research teams took to identify potential bot participants, outline strategies to mitigate the bots' effects and reduce the chances of bot attacks in future projects, and provide recommendations for other researchers.

Ways to Identify Bot Responses

Several signs in both case studies indicated likely bot attacks—both in terms of researcher review of the dataset and the survey metadata and through software flagging functions. We also attempted to develop an algorithm to predict the likelihood of the participant being a bot to help determine authentic participants from bots. These strategies are discussed further.

Researcher Strategies to Identify Bots

In both studies, recruitment appeared far more efficient than expected, particularly as there was no targeted advertising and the recruitment methods used were broad. Both studies had high numbers of participant registrations in a brief period, with registrations frequently being close together (often less than minutes apart).

Second, study participant demographics were disproportionate to the sample pool, and populations often hard to recruit were overrepresented. As this was more subtle than the high number of registrations, this may not have been detected until data analysis was underway if unusual recruitment patterns were not present. Overrepresentation was particularly obvious for CS2, where we had planned to purposefully recruit based on ethnicity and monitored to ensure we were successful in this recruitment but had yet to start targeted recruitment. It is noted that although the proportion remained high in the corrected dataset, this was overinflated due to the small sample size and likely would not have remained proportionally high once higher numbers or participants were recruited.

Third, a review of participant email addresses indicated that most suspected bot attacks had emails that followed similar patterns (eg, a combination of letters and numbers followed by the exact email domains) or were nonsensical. While not all email users choose not to use their name in their emails, and the desire to protect against identity theft means that some people do not have emails that make sense, it was the similarity between the email addresses that highlighted the differences between suspected bots and the participants that we believed to be genuine.

Finally, unusual responses in the internet-based surveys also highlighted potential bot responses. Specifically, random answering for survey questions as evidenced by inconsistent or incongruent responses, a lack of answers to qualitative or text responses, and patterned survey responses indicated likely bots.

Use of Survey Platform Data

In terms of using the functionality of the survey platforms, several additional strategies can be used to determine if a bot attack has occurred. Some survey platforms can flag surveys that are likely bots, so this should be used where possible. Although responses are flagged, researchers still need to review the responses as above to identify atypical patterns of responses. While this requires researchers to review survey responses, the flagging system can potentially reduce the time required to identify bot attacks.

Potential Statistical Prediction to Identify Bots

Overview

For CS2, there were two parts to the study; participants were invited (1) to complete the "research" questionnaire, which consisted of 8 forms, and (2) to use the internet-based intervention component of the study accessed through a second external website using their contact details. By cross-validating the information from the intervention platform with the initial research survey data (N=503), we obtained confirmation of true participants (n=27). However, we could not be sure whether the high number of remaining participants were bots or actual participants who did not continue to the next part of the study, although we assumed they were bots.

XSL•FO
RenderX

Yet this scenario presented us with a unique opportunity to attempt to create a bot detection algorithm to distinguish bots from genuine participants in CS2 better using strategies previously recommended in the literature [18,22]. To achieve this, we assigned suspicion scores on a scale of 0-10 to responses that we evaluated as being potentially anomalous based on the following criteria:

Survey Completion Time

We tracked the completion rates, completion times of individual forms within the survey, and the time difference between the initial and final form completion. A weak bimodal distribution appeared for completion times in CS2, allowing us to apply a higher suspicion score to completion times of less than 5 minutes.

Email Address Analysis

We scrutinized patterns in email addresses, such as random strings of letters and digits, along with unusual domain names, to identify suspicious email patterns typically associated with bots. Higher suspicion scores were applied to unusual email address combinations.

Conflicting Response Analysis

We evaluated responses to specific correlated survey questions to detect inconsistencies that might indicate automated responses. For instance, we sought contradictory answers to 2 questions concerning the frequency of experiencing "not being able to stop or control worrying" and "worrying too much about different things" over the past 2 weeks. Greater negative correlations between paired questions were assigned higher suspicion scores.

To construct the bot detection model, we used the extreme gradient boosting (XGBoost; The XGBoost Contributors) algorithm with a logistic regression base model for binary identification of bots and participants. XGBoost is a scalable tree-boosting system that has shown excellent performance in classification tasks due to its proficiency in managing high-dimensional data and capturing intricate patterns [23,24]. We partitioned the dataset, allocating 80% for model training and 20% for testing model performance. Hyperparameters were fine-tuned through experimentation and cross-validation to attain optimal performance. We used receiver operating characteristic curve results to optimize accuracy, sensitivity, and specificity by adjusting the classification threshold. After optimization, the model achieved an overall accuracy of 0.937, with a sensitivity of 0.967 (reflecting a high rate of bot detection), but a specificity of 0.400 (reflecting a poor rate of actual participant detection), for a classification threshold of 0.25. Hence, the optimized model was heavily biased toward detecting potential bots and unable to identify true participants consistently. The model's balanced accuracy was moderately low at 0.683, reflecting the challenge of simultaneously identifying true positives and true negatives.

Based on this model, the most important features in order of predictive use were the time difference between the first and last form completions within the baseline survey, the time difference between adjacent forms from "unique" participants, the total completion time of the overall survey, the email suspicion score, and conflicting responses for correlated questions. We also ran univariate logistic regressions for individual features to validate the XGBoost model results. Heavy 0 inflation (ie, high numbers of participants with suspicion scores of 0) and poor discriminability were present in each feature, making it impossible to identify bots accurately even when features were combined.

To perform supervised learning this way, we assumed a binary classification of participants versus bots. In reality, the data were imperfectly labeled. Instead of labeling participants as "not bot" or "bot," the best we could achieve was labeling them as "not bot" or "maybe bot." Because the study comprised the internet-based research questions and the internet-based intervention, we used the login and intervention usage data to confirm some true participants (as the intervention website required interaction with diary components and activities, thus confirming a true participant). Nevertheless, there remained a high degree of uncertainty regarding the identity of the remaining participants who only completed the first baseline research questionnaire but did not proceed to log into the intervention. Thus, the bot detector's accuracy is probably exaggerated.

Ways to Manage a Bot Attack

Researchers should regularly monitor recruitment and data collection to aid in the early identification of bot attacks. Close monitoring is critical soon after studies are advertised on the web, as the risk of a bot attack, based on these 2 studies, seems to occur soon after study advertisement. While frequent data monitoring will not stop a bot attack, it may reduce the breadth of the attack by being able to intervene early. Researchers can monitor data manually by regularly checking recruitment or by setting up alerts for new enrolments, which will signal if a high volume of enrolments occurs over a short time.

Close or Pause the Survey

Once a bot attack is detected, it is essential to close or pause the survey immediately. In both case studies, new surveys were created and circulated with additional security measures in place, including location screening features, CAPTCHA coding, the use of fraud scores algorithms, and referral restrictions if recruitment occurs through social media (further strategies are discussed in the section "Ways to counteract a bot attack"). Closing or pausing the survey ensures that no new bot enrolments can occur (although settings may allow incomplete surveys to be completed) and extra security measures can be implemented.

Seek Institutional Support

After identifying the bot attack, several actions were taken to help the research teams consider what to do next. Actions included notifying the respective ethics committees about the attack and seeking consultation and advice about responding. Despite this surge in internet use for health research, there remains limited guidance from health research ethics committees on managing bot attacks or ensuring the validity of internet-based research studies. The emergence of large-scale

XSL•FO
RenderX

digital research and artificial intelligence have presented unique challenges in research ethics and safety. Artificial intelligence (AI) and "big data," particularly digital data, present unique ethical considerations, which institutional research ethics boards may not be adequately equipped to deal with [25]. For example, the issue of transparency is unique to contemporary AI models, as they are often considered "black boxes" where the user is unaware of how the model has reached its conclusion [26]. Accountability remains a complex issue as it is unclear who should be held responsible when AI is involved [27]. In this context, we believe retaining autonomy and "human control" of digital data is paramount. AI can also blur the concept of free and informed consent, and managing privacy becomes more challenging with AI's ability to identify individuals even after data deidentification [28]. Data bias is also a concern as AI may not always detect or could potentially generate biased results, including harmful gender and racial biases [29]. It is crucial to continue addressing these ethical challenges to ensure responsible and safe implementation of AI technologies.

In both our case studies, several ethical dilemmas arose. The first dilemma was the issue of reimbursement. As there is often a set budget for participant reimbursement, an ethical dilemma arises when the number of claims exceeds those budgeted. Specifically, in studies where actual enrolments exceed planned enrolments, there could potentially be hundreds of people who cannot be reimbursed if the study protocol outlines reimbursement of all respondents regardless of data quality. The second dilemma was the inability to differentiate between bots and actual participants in a way that could 100% confirm which responses were "real" and which were not. Failure to accurately identify between participants and bots could lead to a risk of reimbursing bots and erroneously excluding reimbursement of actual respondents due to budgeting constraints.

Finally, there were concerns about potential reputational risks for the researchers and their academic institutions if the research team reimbursed "fake" responses or missed payment for "real" participants. The respective university ethical and legal teams were consulted about researcher obligations to compensate based on completed answers, even if suspected bot attacks generated these. The standard advice was that the bot attack was a misrepresentative response, and therefore, there was no obligation to compensate, given that the terms of research participation had not been met.

In CS1, where participants were to be sent vouchers for compensation, careful screening of voucher claims revealed only 4 valid registrations (out of over 1200 claims), which were subsequently paid. Where participant contact details were available in CS2, we gained ethical approval to recontact participants to indicate that there had been an attack on the study. We asked participants to complete their baseline data collection again to obtain their compensation. The recompletion rate was approximately 55% and no complaints were received about this by the researchers or the ethics committee.

Remove Data

Where there was a high probability of being a bot participant, these data were removed from the dataset, and a record was made of this decision. This decision was based on the above identification parameters. Where it was difficult to tell if the participant was a bot-participant or a legitimate participant, this was noted to allow further scrutiny during the longitudinal follow-up and for reporting in publications as appropriate. It is noted that in both case studies, 2 researchers were involved in each review of data and the decision to remove participants based on being a bot. It was also determined that it was important to do this before data analysis occurred, document all research decisions, and inform the respective ethics committees.

Ways to Counteract Bot Attacks

Plan Ahead

We believe that all researchers involved in internet-based research should have a bot-management protocol in their data management plan. This could be based on protocols such as Ballard et al [11] and include the frequency of monitoring enrolments for bot attacks, steps to take when the survey is active if a bot attack is suspected, and a data handling and analysis plan for the study and that includes planning for bot-related and suspicious data. There should be a clear justification for removing data, and the dataset should be reviewed independently by 2 reviewers, including any decision made to remove data. Decisions should be documented, and publications related to the study should disclose the level of data removal related to the bot attack. Similarly, ethics committees who have approved the study should be consulted and informed about bot attacks, given the impact on research integrity.

Study Advertisements

The researchers in this study and others [9,18,30] believe that studies that are advertised on social media, particularly media where bots are highly active and reach are widespread (such as Twitter or Facebook), are more likely to experience bot attacks. The placement of study advertisements should be specific to the study population, where possible, and highlight the benefits of participating in the research other than compensation. Settings on survey platforms can also limit access to pages via routes other than where they were initially posted; for example, in CS1, settings were changed so that only participants who accessed the survey via the Facebook pages where the links were posted were included in the dataset. In addition, only those who accessed the voucher registration page via the survey could enter their details. Using previously verified distribution lists, such as those associated with professional or patient organizations, may reduce the likelihood of bot attacks as the advertising is more targeted.

Participant Information Sheets and Consent Forms

Participant information sheets and consent forms should highlight the steps researchers will take if they believe a participant has misrepresented themselves. Currently, participant information sheets often outline the inclusion and exclusion criteria and may ask that the participant indicate that they meet these criteria; however, there is no explicit statement about data management for suspected bot attacks. Upon reviewing the

XSL•FO
RenderX

participant information sheet for CS2, the ethics committee felt that the wording and context did sufficiently imply that reimbursement was given only if the questionnaires were completed as part of the intervention study, so this provided adequate grounds for researchers to omit reimbursement to participants who did not complete the questionnaires as part of the fuller study. Nevertheless, an ethics amendment was made where the wording was changed to stipulate that reimbursement more clearly was only given if the questionnaires were completed and the study intervention was used as part of participation.

For future studies, we would recommend the inclusion of text such as the following:

> *Any data that is thought to be generated by a bot or non-human means, is a duplication beyond what is required by the study, or is misrepresentative may be removed by the research team. If this occurs, study compensation for participation, as outlined in this participant information sheet, will not be offered. If you believe your data has been removed unfairly, you may discuss this with the lead researcher, who may ask you for proof of the legitimacy of your claim. If there are issues with data collection, you may be contacted to complete parts of the research again, which you may choose to decline.*

Consent forms should also include reference to this, such as "I understand that if my data is believed to be misrepresented it may be removed from the study and I may not receive compensation for participation." Before including this in external documents, we recommend consulting with your institutional legal team and ethics committee.

Prescreening of Participants

Designing a 2-step registration process that requires screening for study eligibility prior to accessing the survey and that includes multifactor authentication for identity validation may be 1 way to counter web-crawler bots. A more thorough strategy could be prescreening participants manually via a phone call, but this increases the time intensity of a study [21] and may also create a hurdle for participation by requiring additional steps to be completed. Further, commercial tools for identity verification from public records may also be used depending on the location of the study, ethics committee approval, and the researcher's ability to access these. However, these tools rely on up-to-date public records and prevent anonymous participation, which again may create barriers.

Compensation of Participants

Offering financial compensation for research participation is a common strategy when engaging in health-related research, encouraging a breadth of responses [31] and helping to reduce the impact of self-selection bias [32]. Compensating participants is considered an ethical approach to ensuring widespread health research engagement and facilitates participation by financially disadvantaged participants [33,34]. Compensation also recognizes the time participants have contributed to the research project and the value of their contribution. However, compensation means that research can also attract participants who may not be interested in the outcome of the research and are participating for other reasons.

Advertising compensation in study advertisements may increase the likelihood of studies being targeted by bots [5,6,18]. Likewise, direct compensation for internet-based data collection increases the risk of a bot attack, which may be mitigated by offering a prize draw instead. However, a prize draw may be less effective for recruitment [35], may not fairly compensate participants for their time, and may be less likely to recruit or may greater disadvantage more some individuals [36]. Alternatively, offering compensation contingent on completing specific tasks may reduce the likelihood of fraudulent participation, particularly if compensation payment is only made at the end of the study.

Ethics committees recommend offering compensation relative to the degree of involvement of the participant. For example, a 5-minute survey may only be eligible for a prize draw, whereas a study with an in-person 3-hour visit that requires blood tests and medical scans will warrant a higher compensation. However, there is limited guidance on what compensation is expected for different activities, and research teams arbitrarily determine compensation. Thus, studies where researchers offer a higher level of compensation for less effort (such as a short internet-based survey), might be more appealing to bot attacks. Given this, we would advocate for more explicit guidance on research compensation, particularly for internet-based studies.

Clarifying participant identity to claim compensation may also be another barrier for financially motivated bot attacks. Examples of this could be participants being asked to provide identification at the end of the survey to claim compensation and to ensure the participant is a person or the research teams sending vouchers to a postal address rather than an email address. Another potential way to determine authenticity may be using electronic bank transfers to a confirmed bank account rather than digital vouchers similar to case transfer programs that have been trailed in health research [37]. Alternatively, cryptocurrency could also be effective, as the process of claiming blockchain-based rewards could involve complex cryptographic tasks or multifactor authentication methods that are easy for humans but difficult for bots. Cryptocurrency transactions are also publicly verifiable by default on the blockchain, providing a transparent record of each payment (note that cryptocurrency has not yet achieved sufficiently widespread adoption to enable this method, and its use could potentially be seen as contentious [38,39]). Likewise, the requirement to share personal banking information may be a barrier to participation or could preclude people who do not have the financial fluidity to experiment with or access cryptocurrency, which has often been associated with people with more financial freedom [40]. Thus, the use of cryptocurrency may only be viable in limited situations. These strategies could reduce the potential for bots or hybrid bot attacks to be paid and could prevent participants from claiming multiple compensations for multiple completions of the same survey.

XSL•FO
RenderX

Enhanced Security Settings

Researchers should familiarize themselves with the security options within their survey platforms to reduce bot attacks. Security options include activating settings to reduce multiple responses from the same participant, bot detection features, and CAPTCHA settings. For CAPTCHA, Qualtrics (Qualtrics International Inc) generates scores for each participant on a 0 to 1 scale based on several variables. Scores less than 0.5 are likely bots. It was noted that in the case studies above, this was not enough to deter bots but was a method of helping to identify potential bots. RelevantID settings also use code to determine if participants are taking the survey multiple times, allowing researchers to remove duplicate participants.

Crafty Survey Development

Two survey design steps may help make it more difficult for bots to access surveys. The first requires registration with an email address, and then potential participants must validate their email address to log on to the internet-based survey. This registration requires more actions to log in, which makes it harder for less sophisticated bots to access the survey. However, as noted in CS2, bots or hybrid bots may be able to overcome this. Additionally, initial screening or inclusion criteria could be set up as individual questions that are randomized in order, and the correct answers need to be selected for inclusion in the final dataset. An alternative could be using questions that require 2 answers that are linked and require only certain answers to indicate a valid participant. An example could include selecting a country of origin from a drop-down list and then entering a postcode [20] which the research team check to validate if this is a real participant.

Surveys can also include questions that are only likely to be completed by bots, not participants. These are called “honey pot” questions, and each survey platform has a different way of creating them. For example, a field may be created that is invisible to humans, but can be detected by a bot. It is also noted that creating these questions in some coding software may be easier for a bot to detect, and recent internet-based threads have recommended that JavaScript (Mozilla Foundation) be the preferred coding platform for creating these fields.

The inclusion of text-based questions that need answering may also help with the identification of bot participants. Examples of this could include a series of questions requiring a typed answer about the country’s capital, what year it is, the city the person lives in, and basic mathematical equations such as “What is one plus three?” Question format can also increase the chance of detecting a bot by directing participants only to select a specific number of items from a list where all items could be answered.

Kay and Saucier [41] have created an internet-based database of 660 items that can be used to detect careless or insufficient-effort responders in survey data. The authors propose that the best way to identify participants who respond in a misleading or invalid way is to add frequency and infrequency items. Frequency items should be endorsed almost by every respondent (eg, it should be illegal to kill an innocent person), and infrequency item statements should be endorsed by almost no respondents (eg, when someone tells a funny joke, I feel angry) [41]. Thus, failure to endorse a frequency item or an infrequency item would likely indicate a bot.

The final recommendation is randomization of survey items or questions that make it difficult for bots to “predict” the question and answer “appropriately,” thus making it easier to identify nonsensical surveys. Similarly, including free text or open text questions may also help to identify bots versus actual participants, as bots are less likely to complete these meaningfully or may replicate answers. Thus, reviewing qualitative answers for quality or content duplication may help identify a bot. Additionally, including qualitative answers may make bot responding more challenging and provide more insights for future research not captured in closed-text answers [14]. It is noted that these strategies individually will not necessarily prevent hybrid people-bot attacks but will increase the difficulty of survey completion. If multiple strategies are used, it will likely be more deterring for the bot and make it easier to identify bot-responding.

Metadata

The use of IP address tracking may be helpful to identify actual participants compared to bots. Coupled with location filters, IP address tracking can help determine when multiple attempts are made of the same survey and when the participant is not within the geographical catchment of the survey. However, it is noted that virtual private networks can circumnavigate this, and IP address tracking alone is not foolproof. Some ethics committees consider IP addresses to be identifiable information that must be disclosed during the informed consent process if collected. Using this strategy should be approved by the overseeing ethics committee. For further details on using IP addresses to mitigate bots’ effects, see White and Brodhead [8].

Asking participants to provide information that is matched to their IP address or metadata could also be another strategy to validate real participants. This could include strategies such as asking what country the person is in and matching it to an IP address or geolocation or to confirm the time where they are and match it to survey functions which capture the times of both the participant’s browser and the server of the host institution to detect discrepancies [20].

Challenges

Overview

There is likely no foolproof way to detect bots and determined bot users could overcome many of these strategies through a human answering some of the screening questions and then running the bot script. Thus, bot-attacks in internet-based research are an ongoing problem to manage. There are also several ethical issues when identifying potential bots.

Decreased Anonymity

Requiring participants to provide a form of identification prevents researchers from collecting anonymous responses. This may be particularly problematic for studies where participants may be reluctant to participate due to fear, stigma, or where the behavior being studied is illegal [42]. There are strategies to

XSL•FO
RenderX

reduce the impact of the potential to be identified, such as the software randomly allocating codes that are not tied to answers or a researcher who does not have access to the survey data (or even the topic) being the contact person for voucher claims. Researchers working in areas where stigma and shame are common may be reluctant to design surveys where participants may need to be identified; however, there is research showing that larger incentives and greater privacy do not necessarily equate to disclosure of more sensitive data [43] and greater anonymity may impact the accuracy of internet-based surveys [44]. Thus, protecting participant identity through deidentification may be the preferred route to prevent bots and increase survey accuracy.

Removal of Data

Removing bot data is important for research integrity, but this should be done cautiously so as not to remove data belonging to valid participants. This is particularly valid when considering removing data with incomplete answers (particularly qualitative answers), as this may result in removing data from people who may struggle with literacy but also misrepresent attrition and engagement—both critical issues in digital interventions [45-47].

Data removal should be done by 2 independent, blinded researchers and disclosed in publications to ensure transparency and rigor and decrease biases. Similarly, sharing the processes around managing bot-suspected material will aid the research community in developing a consistent approach to bot-data management. Specifically, we believe that researchers should report the methods used to determine bot data, the percentage of data removed due to the data being bot-generated, and the percentage of any suspicious remaining data. As such, we need consistent community guidelines for handling and reporting bot-related data and research impact.

Abusive Responses or Researcher Protection

Soon after the early closure of one of the studies, the researchers began to receive requests for participation compensation. When "participants" were told that there would be no compensation due to suspected fraudulent behavior, researchers began to receive abusive messages. While this is often not at the level of abuse that has recently been disclosed with the rise of the internet [48,49], receiving such abuse is a risk to the well-being of researchers [50,51]. Given this, a plan is needed to determine if and how to respond to such communications and ensure the researcher's well-being [50,51].

Discussion

Principal Findings

This study presents 2 case studies of bot attacks encountered by our research team and the strategies we used to counter them. Moreover, it underscores the critical challenge facing internet-based survey research, which remains highly susceptible to various forms of malicious interference. Unless novel technologies are developed to counter these threats, the challenge of obtaining trustworthy survey data over the internet will likely become more difficult in future years. Our hope is that this paper can serve as a warning, helping to drive the changes needed to protect the integrity of internet-based survey research moving forward.

As technology evolves, so too will the sophistication of bot attacks. As such, bots are here to stay. Therefore, researchers will need multiple, regularly updated strategies to combat and manage bots' effects, including institutional support and ethics committees knowledgeable in this space. It is likely that health research will need to consider multiple bot prevention strategies, use multimodal recruitment and data collection, and develop clear guidelines to help researchers manage bot-related data in internet-based research.

Authentic participant data are valuable, and therefore, researchers need algorithms and a decisional process to detect bot data. A method that is overly sensitive to bots may remove actual participants unnecessarily, thus reducing the size of the dataset and may result in the study being underpowered. On the other hand, an algorithm or decisional process with low specificity may allow an excessive number of bot data to be included, reducing data quality and increasing study costs through compensation paid to nonauthentic participants. Further, bots are likely to provide either extremely noisy or biased answers that negatively impact data quality. When creating a bot detection algorithm, it is up to the researcher to decide whether they wish to prioritize sensitivity (the ability to detect participants) or specificity (the ability to detect bots).

When attempting to develop an appropriate algorithm to detect bots, we faced severe issues with zero-inflated data where most participants—whether bots or not—registered a suspicion score of 0 for multiple variables. This indicates a lack of information about the methods used to create the bot. It highlights the possibility that the attack was performed by human actors responding manually to each survey, making it seem more authentic. Yet even with perfect knowledge of our artificial bot dataset, any bot detection algorithm trained on this dataset would endure from severe overfitting issues, as the attack strategies used will likely differ from those used for another project. Consequently, our detection algorithm probably would not generalize well to real-world attacks, except for similarly designed REDCap (Research Electronic Data Capture; Vanderbilt University) surveys. To make this even more challenging, each of the survey tools currently available to researchers (such as Qualtrics, REDCap, or SurveyMonkey) have different programming interfaces and database structures that would each require a unique version of the bot detection algorithm that is configured to their setup. Hence, we remain unaware of any tool that provides sufficient accuracy for bot detection across the wide range of currently used survey tools. Detection is made more difficult by imperfect participant labeling. For example, researchers can rarely be completely certain that an email address belongs to a bot, even in the presence of "obvious" signs such as random strings of letters or digits. Genuine participants wishing to preserve their anonymity on the web may create a dummy email address for responding to surveys or use "hide my email address" systems, which can also create unusual email handles. The best way to counter fake email addresses is by screening potential participants beforehand and sending a private, individual survey link to prescreened addresses rather than making the link

XSL•FO
RenderX

publicly available [18]. However, this may not be practical for all projects as it can be labor-intensive and as highlighted in CS2, is not foolproof.

Implications

Health research often engages vulnerable populations, and special care must be taken to protect privacy and ensure appropriate risk management is available where needed—including the detection of bots that may skew findings. One vulnerable group are the Indigenous populations. Health research must be responsive and capture the needs of indigenous populations who often experience poorer health outcomes. In CS1, bot responses identified as Māori (the Indigenous population of New Zealand) made up 30% of the sample, and after bot measures were in place, the percentage identifying as Māori was 10%. In CS2, "bot" responses inflated the participants that identified as Māori to almost 4 times what we expected in recruitment based on the population composition. Had we not identified the bot, we would have been quite confident in the analysis we made relating to Māori participants, given the high number, despite these results being unrelated to Māori at all. Therefore, there is a risk that failure to detect and adapt to bot technology could lead to misinformation and contribute to poorer outcomes for indigenous populations based on misinformed conclusions.

Given this, we encourage researchers to consider whether internet-based surveys are the best way to obtain a representative sample with high data accuracy. Internet-based recruitment may preclude some people, and although we increasingly see people being connected on the web, regular data connection, access to technology, and technology literacy may not yet be everyone's privilege, including vulnerable communities. This is particularly pertinent for those where the digital world may be at odds with core and cultural values [52] or where surveys may not accurately capture experiences [53]. Instead, the use of multimodal surveys may reach a broader group, including those who may not be aware of internet-based surveys, may be more responsive to the needs of communities, and may be able better to manage the impact of bots [14].

Recommendations

From our experiences and literature review, it would seem prudent to develop clear guidelines for conducting internet-based research to reduce the risk of bot attacks and increase the robustness of internet-based research. There are currently few guidelines that exist to guide practice in this area. The Association of Internet Researchers released their latest ethical guidelines for internet research in 2019 [54]. The guidelines provide useful information about data management and security, and consider some of the critical issues that we discuss in our paper, including how to protect the researcher where the researcher's public identity is known; specific ethical topics such as accountability, trust, and transparency which have different considerations for internet-based research; and issues related to the accuracy of data including in-built biases from algorithms used for collection and how to use metadata. The EQUATOR (Enhancing the Quality and Transparency of Health Research) network refers to 3 reporting guidelines relating to digital health research—the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist—a checklist for reporting results of internet e-surveys [55]; the CONSORT (Consolidated Standards of Reporting Trials)-EHEALTH to standardize the reporting of evaluations of web and mobile interventions [56], and the iCHECK-DH (Guidelines and Checklist for the Reporting on Digital Health Implementations) which is a guideline and checklist on how to report digital health implementations [57,58]. However, these EQUATOR guidelines refer more to the reporting of the internet-based study than the conduct and do not refer to bot attacks or management [59]. We advocate for developing standardized guidance for researchers conducting internet-based research that describe key considerations and a standardized approach to ensuring data accuracy and validity. These guidelines should also include recommendations for reporting when bot attacks may have occurred and how researchers should handle data with regard to this. While we recognize that bot attacks are likely to be ever-evolving and are likely to outpace the development of guidelines, we believe it would still be useful for guidelines or discussion documents for researchers for strategies to mitigate or minimize the potential effects of bots. Similarly, these guidelines or discussion documents may include points to help researchers determine if the strategies required to minimize bot attacks and the probability of obtaining valid data outweighs the risk of an attack, thus determining if internet-based health research is viable.

Future Directions

Bot attacks will likely become more widespread and more difficult for researchers to detect as botware and attack algorithms become more sophisticated. The experiences of our research group and others [5,13,17-19] highlight a critical, growing problem that deserves more focused researcher attention. Nevertheless, internet-based research has many advantages, such as the ability to reach large numbers of participants and the ability to complete the research remotely. Our experiences should not deter researchers from conducting internet-based research studies. Instead, our paper is a call to action to raise awareness and encourage researchers to consider the risks and benefits of internet-based research. We recommend developing guidelines around detecting and managing bot data in internet-based surveys to help raise awareness of these issues, provide guidance around survey design and data management, and encourage transparency in reporting data that bots may have impacted.

Researchers should also consider the recent exponential development of large language models (LLMs) such as ChatGPT (OpenAI) [60]. Unlike manually programmed bots, LLM-assisted bots can interpret and respond to surveys more coherently, making LLM responses more challenging to detect. This also allows attackers to automate responses to qualitative questions, a task previously reliant on human guidance. We believe it is only a matter of time before LLM-assisted bots become sophisticated enough to respond to any survey design. Consequently, it may become impossible to trust any results obtained from public survey links. Therefore, we recommend that researchers begin implementing more effective security strategies if they have not already done so.

XSL•FO
RenderX

A question that arises at this point, and a potential focus for further research, is whether bot attacks are generalizable across different parts of the world, as different countries have varying protocols for paying research participants [61]. Our experience also suggests that all current research platforms may be vulnerable to illegal bot attacks, and the associated responsibility of software developers to ensure the security and privacy of people is paramount. Thus, bot attacks are likely to be a global issue.

Conclusions

While internet-based research studies increase the ease of participant recruitment and accessibility to a diverse range of respondents, the rise of sophisticated bot programmers and algorithms to automate survey responses risks invalidating internet-based research. Careful planning of internet-based research study designs and incorporating measures to minimize bot responses, such as using a mix of closed, open-ended, and randomized questions, is necessary to protect internet-based studies from bot attacks. However, these measures should be weighed against the risk of inadvertently disqualifying or turning away real, genuine participants. There is an urgent need for standardized practices and guidelines to be developed to provide researchers with clear guidance on safeguarding against bot attacks and actions to take if a bot attack is suspected. As bot attacks are here to stay, this paper aims to raise researchers' awareness and create a call to action before the problem becomes more widespread and challenging to manage.

Authors' Contributions

LD, NH, MP, AK, and AHYC performed the conceptualization. LD, NH, and HW contributed to the data curation. NH, HW, and LD did the formal analysis. LD, NH, AK, and AHYC did the methodology. LD, NH, AK, and AHYC contributed to the investigation. NH contributed to the software. NH and MP contributed to the validation. LD, NH, AK, HW, and AHYC contributed to writing the original draft. LD, NH, AK, MP, HW, and AHYC contributed to writing—reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Case studies.
[DOCX File , 34 KB - ijmr_v14i1e60548_app1.docx]

References

1. Mills MC, Rahal C. A scientometric review of genome-wide association studies. Commun Biol 2019;2(1):9 [FREE Full text] [doi: 10.1038/s42003-018-0261-x] [Medline: 30623105]
2. van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: an overview. Neuroimage 2013;80:62-79 [FREE Full text] [doi: 10.1016/j.neuroimage.2013.05.041] [Medline: 23684880]
3. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12(3):e1001779 [FREE Full text] [doi: 10.1371/journal.pmed.1001779] [Medline: 25826379]
4. Ward M, Meade AW. Dealing with careless responding in survey data: prevention, identification, and recommended best practices. Annu Rev Psychol 2023;74(1):577-596 [FREE Full text] [doi: 10.1146/annurev-psych-040422-045007] [Medline: 35973734]
5. Griffin M, Martino RJ, LoSchiavo C, Comer-Carruthers C, Krause KD, Stults CB, et al. Ensuring survey research data integrity in the era of internet bots. Qual Quant 2022;56(4):2841-2852 [FREE Full text] [doi: 10.1007/s11135-021-01252-1] [Medline: 34629553]
6. Lawlor J, Thomas C, Guhin AT, Kenyon K, Lerner MD, Drahota A. Suspicious and fraudulent online survey participation: introducing the REAL framework. Methodological Innovations 2021;14(3). [doi: 10.1177/20597991211050467]
7. Dittrich D, Leder F, Werner T. A Case Study in Ethical Decision Making Regarding Remote Mitigation of Botnets. Berlin, Germany: Springer; 2010.
8. White AN, Brodhead MT. Detecting fraudulent responses in online survey research. PsyArXiv Preprints Preprint posted online on May 26, 2023 [FREE Full text] [doi: 10.31234/osf.io/wuk2y]
9. Pozzar R, Hammer MJ, Underhill-Blazey M, Wright AA, Tulsky JA, Hong F, et al. Threats of bots and other bad actors to data quality following research participant recruitment through social media: cross-sectional questionnaire. J Med Internet Res 2020;22(10):e23021 [FREE Full text] [doi: 10.2196/23021] [Medline: 33026360]
10. Smyk M, Tyrowicz J, van der Velde L. A cautionary note on the reliability of the online survey data: the case of wage indicator. Sociol Methods Res 2018;50(1):429-464. [doi: 10.1177/0049124118782538]
11. Ballard AM, Cardwell T, Young AM. Fraud detection protocol for web-based research among men who have sex with men: development and descriptive evaluation. JMIR Public Health Surveill 2019;5(1):e12344 [FREE Full text] [doi: 10.2196/12344] [Medline: 30714944]

XSL•FO
RenderX

12. Dupuis M, Meier E, Cuneo F. Detecting computer-generated random responding in questionnaire-based data: a comparison of seven indices. Behav Res Methods 2019;51(5):2228-2237. [doi: 10.3758/s13428-018-1103-y] [Medline: 30091086]
13. Bybee S, Cloyes K, Ellington L, Baucom B, Supiano K, Mooney K. Bots and nots: safeguarding online survey research with underrepresented and diverse populations. Psychol Sex 2022;13(4):901-911 [FREE Full text] [doi: 10.1080/19419899.2021.1936617] [Medline: 36439051]
14. Yazici E, Wang Y. Attack the bot: Mode effects and the challenges of conducting a mixed-mode household survey during the Covid-19 pandemic. Int J Soc Res Methodol 2023;27(6):791-796. [doi: 10.1080/13645579.2023.2241797]
15. Barrett NM, Tamatea J, Morgan R, Jones AR. Creating an environment to inform, build, and sustain a Maori health research workforce. J R Soc N Z 2023;55(1):47-61. [doi: 10.1080/03036758.2023.2235303]
16. Corsini-Munt S, Rancourt KM, Dubé J, Rossi M, Rosen N. Vulvodynia: a consideration of clinical and methodological research challenges and recommended solutions. JPR 2017;10:2425-2436. [doi: 10.2147/jpr.s126259]
17. Loebenberg G, Oldham M, Brown J, Dinu L, Michie S, Field M, et al. Bot or not? Detecting and managing participant deception when conducting digital research remotely: case study of a randomized controlled trial. J Med Internet Res 2023;25:e46523 [FREE Full text] [doi: 10.2196/46523] [Medline: 37707943]
18. Storozuk A, Ashley M, Delage V, Maloney EA. Got bots? Practical recommendations to protect online survey data from bot attacks. TQMP 2020;16(5):472-481. [doi: 10.20982/tqmp.16.5.p472]
19. Goodrich B, Fenton M, Penn J, Bovay J, Mountain T. Battling bots: experiences and strategies to mitigate fraudulent responses in online surveys. Appl Econ Perspect Pol 2023;45(2):762-784. [doi: 10.1002/aepp.13353]
20. Wang J, Calderon G, Hager ER, Edwards LV, Berry AA, Liu Y, et al. Identifying and preventing fraudulent responses in online public health surveys: lessons learned during the COVID-19 pandemic. PLOS Glob Public Health 2023;3(8):e0001452 [FREE Full text] [doi: 10.1371/journal.pgph.0001452] [Medline: 37610999]
21. Glazer JV, MacDonnell K, Frederick C, Ingersoll K, Ritterband LM. Liar! Liar! identifying eligibility fraud by applicants in digital health research. Internet Interv 2021;25:100401 [FREE Full text] [doi: 10.1016/j.invent.2021.100401] [Medline: 34094883]
22. Xu Y, Pace S, Kim J, Iachini A, King LB, Harrison T, et al. Threats to online surveys: recognizing, detecting, and preventing survey bots. Soc Work Res 2022;46(4):343-350 [FREE Full text]
23. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artif Intell Rev 2020;54(3):1937-1967. [doi: 10.1007/s10462-020-09896-5]
24. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939785]
25. Bouhouita-Guermech S, Gogognon P, Bélisle-Pipon JC. Specific challenges posed by artificial intelligence in research ethics. Front Artif Intell 2023;6:1149082 [FREE Full text] [doi: 10.3389/frai.2023.1149082] [Medline: 37483869]
26. Brożek B, Furman M, Jakubiec M, Kucharzyk B. The black box problem revisited. real and imaginary challenges for automated legal decision making. Artif Intell Law 2023;32(2):427-440. [doi: 10.1007/s10506-023-09356-9]
27. Sand M, Durán JM, Jongsma KR. Responsibility beyond design: physicians' requirements for ethical medical AI. Bioethics 2022;36(2):162-169 [FREE Full text] [doi: 10.1111/bioe.12887] [Medline: 34089625]
28. Erlich Y, Shor T, Pe'er I, Carmi S. Identity inference of genomic data using long-range familial searches. Science 2018;362(6415):690-694 [FREE Full text] [doi: 10.1126/science.aau4832] [Medline: 30309907]
29. Jain A, Brooks JR, Alford CC, Chang CS, Mueller NM, Umscheid CA, et al. Awareness of racial and ethnic bias and potential solutions to address bias with use of health care algorithms. JAMA Health Forum 2023;4(6):e231197 [FREE Full text] [doi: 10.1001/jamahealthforum.2023.1197] [Medline: 37266959]
30. Orabi M, Mouheb D, Al Aghbari Z, Kamel I. Detection of bots in social media: a systematic review. Inf Process Manag 2020;57(4):102250. [doi: 10.1016/j.ipm.2020.102250]
31. Abdelazeem B, Abbas KS, Amin MA, El-Shahat NA, Malik B, Kalantary A, et al. The effectiveness of incentives for research participation: a systematic review and meta-analysis of randomized controlled trials. PLoS One 2022;17(4):e0267534 [FREE Full text] [doi: 10.1371/journal.pone.0267534] [Medline: 35452488]
32. Kaźmierczak I, Zajenkowska A, Rogoza R, Jonason PK, Ścigała D. Self-selection biases in psychological studies: personality and affective disorders are prevalent among participants. PLoS One 2023;18(3):e0281046 [FREE Full text] [doi: 10.1371/journal.pone.0281046] [Medline: 36888578]
33. Jennings CG, MacDonald TM, Wei L, Brown MJ, McConnachie L, Mackenzie IS. Does offering an incentive payment improve recruitment to clinical trials and increase the proportion of socially deprived and elderly participants? Trials 2015;16(1):80 [FREE Full text] [doi: 10.1186/s13063-015-0582-8] [Medline: 25888477]
34. Permuth-Wey J, Borenstein AR. Financial remuneration for clinical and behavioral research participation: ethical and practical considerations. Ann Epidemiol 2009;19(4):280-285. [doi: 10.1016/j.annepidem.2009.01.004] [Medline: 19230712]
35. Dykema J, Stevenson J, Jones CP, Day B. Guaranteed incentives and prize drawings: effects on participation, data quality, and costs in a web survey of college students on sensitive topics. Soc Sci Comput Rev 2023;42(3):777-793. [doi: 10.1177/08944393231189853]

XSL•FO
RenderX

36. Bierer BE, White SA, Gelinas L, Strauss DH. Fair payment and just benefits to enhance diversity in clinical research. J Clin Transl Sci 2021;5(1):e159 [FREE Full text] [doi: 10.1017/cts.2021.816] [Medline: 34527298]
37. Lagarde M, Haines A, Palmer N. Conditional cash transfers for improving uptake of health interventions in low- and middle-income countries: a systematic review. JAMA 2007;298(16):1900-1910. [doi: 10.1001/jama.298.16.1900] [Medline: 17954541]
38. Valdeolmillos D, Mezquita Y, González-Briones A, Prieto J, Corchado JM. Blockchain Technology: A Review of the Current Challenges of Cryptocurrency. Cham, Switzerland: Springer International Publishing; 2020.
39. Alzahrani S, Daim TU. Analysis of the cryptocurrency adoption decision: literature review. 2019 Presented at: Portland International Conference on Management of Engineering and Technology (PICMET); August 25-29, 2019; Portland, OR. [doi: 10.23919/picmet.2019.8893819]
40. Steinmetz F, von Meduna M, Ante L, Fiedler I. Ownership, uses and perceptions of cryptocurrency: results from a population survey. Technol Forecast Soc Change 2021;173:121073. [doi: 10.1016/j.techfore.2021.121073]
41. Kay CS, Saucier G. The comprehensive infrequency/frequency item repository (CIFR): an online database of items for detecting careless/insufficient-effort responders in survey data. Personal Individual Differ 2023;205:112073. [doi: 10.1016/j.paid.2022.112073]
42. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant. Int J Qual Methods 2015;14(5). [doi: 10.1177/1609406915621420]
43. Murdoch M, Simon AB, Polusny MA, Bangerter AK, Grill JP, Noorbaloochi S, et al. Impact of different privacy conditions and incentives on survey response rate, participant representativeness, and disclosure of sensitive information: a randomized controlled trial. BMC Med Res Methodol 2014;14(1):90 [FREE Full text] [doi: 10.1186/1471-2288-14-90] [Medline: 25027174]
44. Lelkes Y, Krosnick JA, Marx DM, Judd CM, Park B. Complete anonymity compromises the accuracy of self-reports. J Exp Soc Psychol 2012;48(6):1291-1299. [doi: 10.1016/j.jesp.2012.07.002]
45. Borghouts J, Eikey E, Mark G, de Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
46. Lipschitz JM, Pike CK, Hogan TP, Murphy SA, Burdick KE. The engagement problem: a review of engagement with digital mental health interventions and recommendations for a path forward. Curr Treat Options Psychiatry 2023;10(3):119-135 [FREE Full text] [doi: 10.1007/s40501-023-00297-3] [Medline: 38390026]
47. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]
48. Bisaillon J, Mercure C, Villeneuve S, Plante I. Cyberbullying of professors: what measures are in place in universities and what solutions are proposed by victims? Stud Higher Educ 2023;48(11):1639-1650. [doi: 10.1080/03075079.2023.2206431]
49. Doerfler P, Forte A, de Cristofaro E, Stringhini G, Blackburn J, McCoy D. "I'm a professor, which isn't usually a dangerous job": internet-facilitated harassment and its impact on researchers. Proceedings of the ACM on Human-Computer Interaction 2021 Oct 18;5(CSCW2):1-32. [doi: 10.1145/3476082]
50. Gelms B. Social media research and the methodological problem of harassment: foregrounding researcher safety. Comput Compos 2021;59:102626. [doi: 10.1016/j.compcom.2021.102626]
51. Mattheis AA, Kingdon A. Does the institution have a plan for that? Researcher safety and the ethics of institutional responsibility. In: Researching Cybercrimes: Methodologies, Ethics, and Critical Approaches. Cham, Switzerland: Palgrave Macmillan; 2021:457-472.
52. Donkin L, Bidois-Putt M, Wilson H, Hayward P, Chan AHY. An exploration of the goodness of fit of web-based tools for māori: qualitative study using interviews and focus groups. JMIR Form Res 2024;8:e50385 [FREE Full text] [doi: 10.2196/50385] [Medline: 38696236]
53. Holcombe-James I. Survey-based research in remote Indigenous communities: considerations for methods. In: Rennie E, editor. Field Guide to Intercultural Research. Cheltenham, United Kingdom: Edward Elgar Publishing; 2021.
54. Heise AHH. Internet Research: Ethical Guidelines 3.0. WildApricot. 2019. URL: https://aoir.org/reports/ethics3.pdf [accessed 2019-10-04]
55. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
56. Eysenbach GE. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
57. Perrin Franck C, Babington-Ashaye A, Dietrich D, Bediang G, Veltsos P, Gupta PP, et al. iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations. J Med Internet Res 2023;25:e46694 [FREE Full text] [doi: 10.2196/46694] [Medline: 37163336]
58. Perrin Franck C, Babington-Ashaye A, Dietrich D, Bediang G, Veltsos P, Gupta PP, et al. Correction: iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations. J Med Internet Res 2023;25:e49027 [FREE Full text] [doi: 10.2196/49027] [Medline: 37201181]

XSL•FO
RenderX

59. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. Lancet 2008;371(9619):1149-1150. [doi: 10.1016/s0140-6736(08)60505-x]
60. Kumar S, Musharaf D, Musharaf S, Sagar AK. A Comprehensive Review of the Latest Advancements in Large Generative AI Models. Cham, Switzerland: Springer Nature; 2023.
61. Grady C, Dickert N, Jawetz T, Gensler G, Emanuel E. An analysis of U.S. practices of paying research participants. Contemp Clin Trials 2005;26(3):365-375. [doi: 10.1016/j.cct.2005.02.003] [Medline: 15911470]

Abbreviations

AI: artificial intelligence
CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CONSORT: Consolidated Standards of Reporting Trials
CS: case study
EQUATOR: Enhancing the Quality and Transparency of Health Research
iCHECK-DH: Guidelines and Checklist for the Reporting on Digital Health Implementations
LLM: large language model
REDCap: Research Electronic Data Capture
XGBoost: extreme gradient boosting

Edited by T de Azevedo Cardoso; submitted 15.05.24; peer-reviewed by E Frechman, L Johnson; comments to author 07.08.24; revised version received 29.08.24; accepted 24.09.24; published 14.03.25.

Please cite as:
Donkin L, Henry N, Kercher A, Pedersen M, Wilson H, Chan AHY
Effective Recruitment or Bot Attack? The Challenge of Internet-Based Research Surveys and Recommendations to Reduce Risk and Improve Robustness
Interact J Med Res 2025;14:e60548
URL: https://www.i-jmr.org/2025/1/e60548
doi:10.2196/60548
PMID:

©Liesje Donkin, Nathan Henry, Amy Kercher, Mangor Pedersen, Holly Wilson, Amy Hai Yan Chan. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.03.2025. 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

Digital Health for Australia: Bridging the Rural, Regional, and Remote Health Gap

Shakeel Mahmood[1], PhD; M Mamun Huda[1], PhD; Kedir Yimam Ahmed[1], PhD; Thapa Subash[1], PhD; Feleke Hailemichael Astawesegn[1], PhD; Anayochukwu Edward Anyasodor[1], PhD; Mohammad Ali Moni[1], PhD; Muhammad J A Shiddiky[1], PhD; Utpal K Mondal[1], MPH; Setognal Birara Aychiluhm[1], MPH; Santosh Giri[1], MPH; Allen Ross[1], PhD

Rural Health Research Institute, Charles Sturt University, Orange, Australia

Corresponding Author:
Shakeel Mahmood, PhD
Rural Health Research Institute
Charles Sturt University
346 Leeds Parade
Orange, 2800
Australia
Phone: 61 0416689348
Email: shmahmood@csu.edu.au

Abstract

In rural Australia, recent trends reveal an exponential increase in the rates of physical inactivity, central obesity, metabolic syndrome, and cancer in the population. The limited rural health workforce, which is struggling to meet this growing burden, is boosted by digital technologies such as My Health Record, Cardihab, Healthdirect, and MindSpot, all of which offer opportunities for improved diagnostics, monitoring, and management of chronic diseases. However, implementing proven digital health technologies in rural communities has been challenging on numerous fronts. This perspective aims to (1) highlight the rural health gap and propose a way forward in implementing evidence-based digital health technologies in the rural, regional, and remote communities of Australia and (2) guide future rural health policy.

(Interact J Med Res 2025;14:e67460) doi:10.2196/67460

KEYWORDS

rural health inequities; digital health adoption; telehealth services; chronic disease management; socioeconomic disadvantage; health workforce; shortages digital literacy gaps; equitable access to care

Rural Health Gap

Approximately 7 million Australians (28%) and 260,000 (32%) First Nations peoples live in rural, regional, and remote (RRR) areas [1,2] (Figure 1 [3]). These regions encompass more than 1700 small towns with populations under 10,000 people [2]. The leading causes of the rural burden of disease are central obesity, type 2 diabetes, coronary heart disease, suicide and self-harm, chronic obstructive pulmonary disease, and lung cancer [4]. Cardiovascular disease is a significant health problem in RRR regions in Australia, where prevalence rates are approximately 20% to 30% higher than in urban areas [4]. Other significant health challenges across RRR communities include chronic kidney disease and dementia and other mental health disorders [5]. As a result, individuals residing in RRR regions experience an age-adjusted mortality rate that is 2.3 times higher than that of those living in major cities [6]. This viewpoint aims to guide policymakers on how to leverage digital health, using global evidence and strategies, to reduce RRR region health inequities.

Several social determinants of health contribute to health disparities in RRR areas, including chronic unemployment, lower educational attainment, historical social injustices, and limited access to fresh fruit and vegetables [7]. For instance, the year 12 completion rate is the lowest in very remote areas compared with that of major cities (55.5% vs 79.4%) [1,8]. Similarly, the median weekly income declines with remoteness, from AUD $1915.2 (the conversion rate at the time of the study was AUD $1=US $0.65) in major cities to AUD $1362.7 in very remote areas [1,8]. The availability of health care professionals also varies substantially, with very remote areas having approximately 205 full-time equivalent doctors per 100,000 people, compared with 427 full-time equivalent doctors per 100,000 in major cities (Table 1) [1,8]. Medicare data from

XSL•FO
RenderX

2022 to 2023 show that general practitioner (GP) visits were lowest in remote and very remote communities, while hospitalization rates (eg, for end-stage renal disease) were almost twice as high in very remote areas and 1.3 times higher in remote areas, with no improvement since 2013 to 2014 [9].

Figure 1. Australian map by remoteness. Adapted from [3].

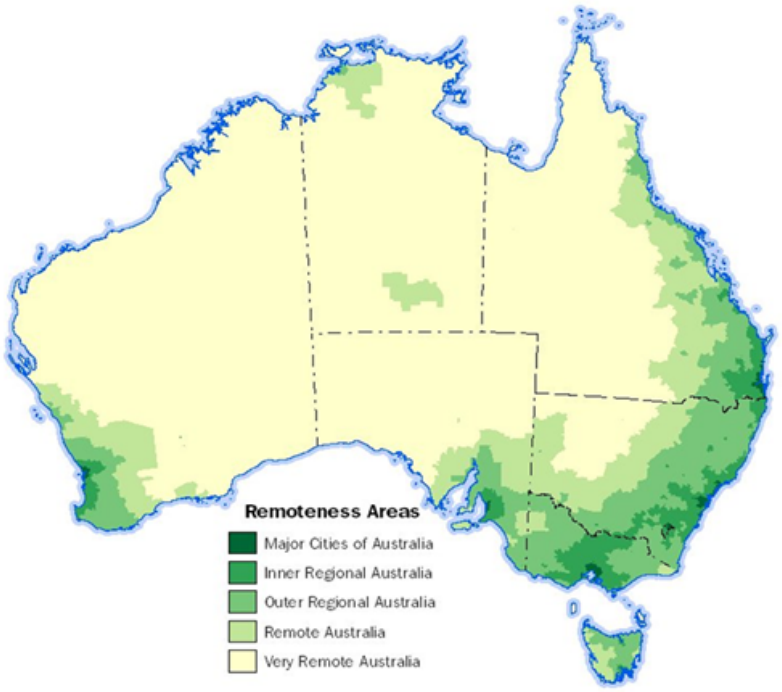


Table 1. Sociodemographic and health workforce profile of Australia by remoteness [1,8].

Variables	Major cities	Inner regional	Outer regional	Remote	Very remote
Population size (%)	72	18	8.1	1.2	0.8
Year 12 completion (%)	79.4	74.5	67.6	69.2	55.5
Median weekly household income (AUD $)	1915.2	1637.6	1391.0	1478.7	1362.7
Health care workforce (medical practitioners), FTE[a] per 100,000	427	311	285	285	205

[a]FTE: full-time equivalent.

Evidence-Based Digital Health Technology

Health care is increasingly being delivered in a technology-driven environment [10]. In Australia, digital health technologies are transforming health care delivery, with significant advancements in electronic medical records (EMRs), telehealth, and remote patient monitoring [11]. EMRs have been widely adopted within GP practices, hospitals, and other service providers, facilitating seamless transfer of patient records among health care providers. Studies indicate that EMR adoption rates in RRR areas often surpass those in urban settings. For example, a 2012 study found that 56% of RRR region practices had adopted EMRs compared with 49% in urban areas [12]. Similarly, another study reported that solo practices in RRR regions were more likely to adopt EMRs than their urban counterparts (41% vs 33%) [13]. My Health Record, a Medicare-funded national digital system, improves care continuity by providing comprehensive, accessible patient information across health care settings (Figure 2) [14].

Telehealth services have also grown substantially, particularly with Healthdirect, a national service funded by federal, state, and territory governments [11]. Healthdirect operates as a "virtual front door," triaging patients, providing self-care advice, and connecting them to necessary health services. Several regions, including Victoria's Northern Health, have implemented "Virtual Emergency Departments" using telehealth to triage patients efficiently [15]. The use of telehealth has extended to innovative outpatient care models, such as the Royal Prince Alfred Hospital's "Spirituals." In the Aboriginal Community Controlled Health Organisation sector, telehealth supports culturally appropriate health care through methods like "Store-and-Forward." Specialized imaging captures patient data and forwards them to specialists for diagnosis and treatment, supporting complex conditions like heart, eye, ear, and diabetes care [16].

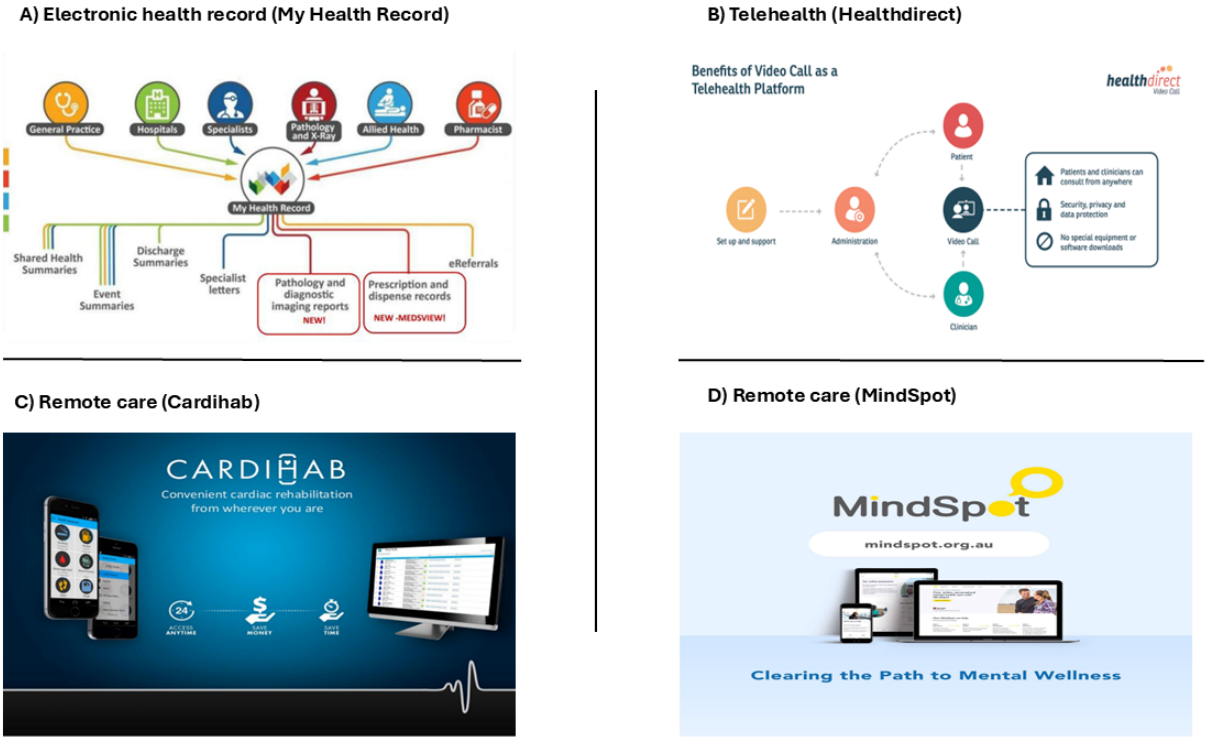
Remote care technologies such as remote patient monitoring and digital therapeutics help in effectively managing chronic conditions and reducing hospital admissions in RRR settings [11]. For example, Cardihab delivers personalized digital cardiac

XSL•FO
RenderX

rehabilitation, enabling clinicians to monitor patient progress through an online portal [17]. Mental health digital therapeutics, provided by services like Mental Health Online, MindSpot, and THIS WAY UP offer stand-alone or blended care options [18]. Blended care enhances flexibility, supports RRR access, and integrates digital tools into traditional pathways for cardiac and mental health services [19]. Hospitals also leverage remote patient monitoring, with the Royal Prince Alfred Hospital's "Virtual Hospital" providing 24/7 multidisciplinary care [20]. The Virtual Clinical Care program in South Australia is another example, aimed at reducing preventable hospital admissions through early intervention and continuous monitoring [21].

Artificial intelligence (AI) integration in health care is still limited, mostly restricted to research settings rather than widespread clinical use [22]. The health sector is one of the least mature industries in terms of AI implementation both in Australia and internationally [23,24]. AI analyzes data to support decisions; telehealth delivers care remotely. Together, they enhance access, efficiency, and personalization [25].

Figure 2. Four available digital platforms in Australia: A) My Health Record securely stores key health information, accessible anytime by patients and healthcare providers, including during emergencies; B) healthdirct operates a 'virtual front door' service. It triages patients, provides advice on how they can manage their condition on their own ('self-care'), and connects them to health services; C) Cardihab is a digital-based cardiac rehabilitation service for patients recovering from a cardiac event or procedure; D) MindSpot is free, confidential psychological assessments and treatments, and access to qualified therapists via online or telephone.



Implementing My Health Record, Healthdirect, Cardihab, and MindSpot

In 2022, nearly half of RRR Australian regions experienced significant broadband or mobile internet connectivity gaps [6]. RRR Australians face ongoing challenges with internet connectivity, especially in terms of video connectivity, and limited digital literacy, despite the narrowing of the urban-rural digital access gap [26]. Implementing My Health Record, an electronic health record (EHR) system, has been challenging on many fronts (Table 2). Issues relate to usability, safety, and the burden it places on physicians [27]. Other challenges of EHR implementation include data entry issues, inadequate alerting systems, and problems with interoperability [27]. Other concerns include high implementation costs, EHR standardization issues, resistance to change, and initial system use reducing productivity [28]. One survey found that most people (71.2%) knew little about My Health Record, and 29.4% had opted out of the system [29]. My Health Record improved access to information, but health literacy gaps, language barriers, and medical jargon caused misinterpretation and sharing issues. Studies need to assess whether the My Health Record improved Australia's health care as per the objectives of the 2012 Act [30].

XSL•FO
RenderX

Table 2. Resources required for the 4 digital health platforms in Australia [17,31-33].

Resources required	My Health Record (electronic health record)	Healthdirect (telehealth)	Cardihab (remote care)	MindSpot (remote care)
Internet	✓	✓	✓	✓
Broadband connections	✓	✓	✓	✓
Electricity	✓	✓	✓	✓
Cellphone				
Normal phone		✓		✓
Smart phone	✓	✓	✓	✓
Computer or tablet	✓	✓	✓	✓
Trained staff	✓	✓	✓	✓
IT staff	✓	✓	✓	✓
Health staff	✓	✓	✓	✓

Healthdirect Australia offers nurse-led telephone triage, crucially supporting RRR areas facing GP shortages and limited health care access. The service helps patients determine the most appropriate level of care, reducing unnecessary emergency department visits [34]. Healthdirect specializes in providing virtual health information, advice, and referral. However, the implementation of such digital health services can be challenging. A key barrier to implementing is the complexity of navigating Australia's health care system [35], where coordination between primary and specialist care is poor, especially affecting access in underserved and RRR communities, according to the Grattan Institute report [36]. This leads to missed follow-ups and fragmented care, especially for patients with chronic or complex conditions [36]. Failure to involve a patient's GP in all aspects of patient care can disrupt continuity of care and compromise the provision of high-quality, cost-effective care [37], particularly in Australia, where the health care system heavily relies on GPs for primary care. A theoretically based implementation plan that can be reported transparently is another pressing requirement [38]. Furthermore, allocating resources to implement the intervention within routine care can also pose difficulties [38].

Cardihab offers digital cardiac rehabilitation [38], improving access and participation in RRR areas, where travel distances historically limited program engagement. In North Queensland, inadequate discharge planning and referrals limit cardiac rehabilitation access for hospital-treated patients [39]. Cost-effectiveness shows savings via fewer rehospitalizations, but sustainability requires continuous Medicare funding and digital skills training for clinicians and patients [40,41]. Despite these benefits, integrating digital cardiac rehabilitation into routine health care remains challenging. Service underuse persists due to a combination of patient and provider factors. Barriers include high costs, staff resistance, training demands, usability issues, data privacy, technical limitations, migration difficulties, and interoperability challenges [38]. National data show that among the 49,900 eligible patients between 2013 and 2015, only 30% were referred to cardiac rehabilitation, and of those referred, fewer than one-third (28%) attended the program [42].

MindSpot delivers online therapy to underserved RRR Australians [43], effectively reducing depression and anxiety symptoms comparable to face-to-face psychological care [44]. Despite these benefits, uptake remains disproportionately low; only 18.9% of registrants are from RRR areas, where approximately 29% of Australians live [45]. Barriers include persistent digital access issues, cultural acceptability concerns, and limited targeted promotion in Indigenous and farming communities [46]. Furthermore, evidence is lacking about the effectiveness of digitally delivered mental health care in real-world settings [45]. Challenges related to the collection of outcome data also persist [47], highlighting the complexities of implementing digital health services in RRR settings.

Additional Examples of Telehealth and AI in RRR Settings

Telehealth is also being leveraged beyond these 4 platforms. For example, the Royal Flying Doctor Service delivers telehealth consultations to remote patients for chronic disease management and mental health, reducing emergency retrievals [48]. The Northern Territory Health Department outlines a structured approach to early detection and management of type 2 diabetes in Aboriginal youth [49]. This includes referral pathways and screening protocols that align with telehealth-enabled services, such as diabetic obesity screening [49]. In addition, Healthy Rural Hearts telehealth in NSW delivers personalized nutrition therapy for cardiovascular disease risk, although its effectiveness and patient engagement are still under evaluation [50].

A systematic review finds evidence of the promising potential of portable ultrasound in RRR Australia, enabling nurses to capture scans and receive automated reports for obstetric and cardiac diagnostics [51]. For example, studies in remote Indigenous communities have shown feasibility and improved diagnostic capacity when nurses use portable ultrasound devices under tele-supervision from specialists [51]. This approach reduces the need for patient travel, supports timely clinical decision-making, and strengthens local health care capacity in under resourced areas. AI-supported triage tools in Healthdirect call systems also show promise, as evaluated in a 2025 Macquarie University usability study of the Triage Platform

XSL•FO
RenderX

[52]. The platform can aid in detecting high-acuity cases but is still in the pilot stages, needing further validation, especially in RRR areas [52].

Current Utilization

Australian residents who are registered with Medicare or an individual health care identifier can access My Health Record via their mobile app or via the official webpage [53,54]. Mental Health Online provides tailored cognitive behavioral therapy programs for anxiety, depression, posttraumatic stress disorder, and obsessive-compulsive disorder, guided by its Electronic Psychological Assessment Screening System [55]. My Health Record is particularly valuable for continuity of care when patients need to travel across large distances for specialist consultations. Rural GPs report improved care coordination, especially for patients with complex multimorbidities, where local facilities cannot provide specialist services [36]. Rural solo practices adopt centralized health records more than metropolitan groups owing to benefits in sparse referral networks [34]. However, full integration is limited because of persistent opt-out rates in remote areas, partly due to mistrust and low awareness [11]. Healthdirect offers 24/7 helplines and online resources, providing Australians with accessible, trusted, and quality health information and advice [54,56]. Cardihab has been used by Australian private health insurers, clinicians, and patients in Queensland, Victoria, and South Australia [57]. Cardihab, a class I medical device registered in Australia, is being implemented nationwide to support cardiac rehabilitation services [58]. In 2023, 27% of MindSpot users were from RRR areas [59]. Mindspot provides mental health services to underserved populations, including Indigenous Australians, those facing access barriers, and residents outside major cities [60]. THIS WAY UP offers evidence-based online mental health courses, supporting RRR region practitioners and providing culturally safe tools for Indigenous communities via the WellMob portal [61].

International Case Studies

In Canada, the Ontario Virtual Care Program, initiated in 2006 and supported by the Ontario Telemedicine Network, introduced virtual visits to assist patients in remote locations [62]. This initiative aimed to enhance access to care and improve continuity and coordination of services. The program has evolved to integrate virtual care services within the health system, benefiting underserved populations, including those in RRR communities [63]. Additionally, the Ontario Virtual Care Program enhances sustainability by enabling policymakers to iteratively adapt service delivery models to RRR realities [64].

In Finland, the OuluHealth ecosystem in northern Finland brings together stakeholders from the health sector to facilitate cooperation and improve health and social care services. Collaboration in medicine, wireless technology, ecosystem enabling, and big data facilitates rapid responses to digital health challenges, effectively addressing societal needs [65]. Integrated digital health models with robust infrastructure can improve health care access and continuity in remote regions.

The Framework for Digital Health Equity study [66] emphasizes that equitable digital health solutions must address barriers for marginalized, Indigenous, and older adults in rural US communities. Ensuring these groups have access to technology and the necessary digital skills will be crucial in reducing health care disparities and promoting digital health equity.

Evaluation Approaches

The evaluation of digital health programs in RRR regions relied on a mixed methods design that incorporates the following:

- Service use data (eg, Medicare claims or platform analytics) to quantify reach and uptake [67]
- Health outcomes (eg, hospital admission rates or disease-specific metrics such as HbA_{1c} also known as glycated hemoglobin levels for diabetes) [68,69]
- Economic evaluations assessing cost-effectiveness considering patient travel, workforce efficiency, and avoided hospitalizations [40]
- Qualitative feedback from patients, clinicians, and community leaders in RRR areas to assess cultural safety, usability [70], and trust

Way Forward

Implementing digital health services in RRR areas is a complex task requiring careful planning and strategic investment (Figure 3 [15,53,57,71,72]) [26]. Rural communities' needs must be assessed to identify current and future essential clinical services [41,73]. This will involve a cross-sectional survey, community consultation, and data analysis [41,73]. Analyzing ethical risks, privacy, security, consent, and regulations can enhance telehealth governance and strengthen digital health policy recommendations [11,74]. Furthermore, systematic data collection would enable benchmarking, monitoring, and ethical oversight, driving continuous improvement across digital health services [75]. Federal and state health authorities must prioritize RRR region–specific data and programs to address unique needs [2]. Once the needs and current capacity are identified, a plan to roll out digital services needs to be developed to complement the existing clinical practice. One such project was the implementation of electronic prescribing infrastructure in 2021, where 36% of GPs generated e-prescriptions, and 95% of the Pharmaceuticals Benefits Scheme pharmacies dispensed them [11]. RRR health service implementation is expensive because of the costs incurred through infrastructure, equipment, skilled workforce, tailored interventions, multilevel strategies, local leadership, and advocacy needs [76,77]. Socioeconomic, connectivity, literacy, and funding gaps worsen RRR region digital health disparities, necessitating inclusive, equity-focused strategies [78-80]. Government subsidies for devices and training programs for RRR health care workers can help reduce barriers and strengthen digital health adoption [81,82]. Government funding; partnerships (policymakers, health care providers, and technology developers) with the private sector; and innovative financing models should be explored to cover these costs [76]. Effective implementation of digital health in RRR areas requires long-term Medicare funding, strengthened public-private partnerships, and targeted workforce training.

XSL•FO
RenderX

Equitable policies, reliable infrastructure, and practical integration of telehealth, EHRs, and remote monitoring are essential. Overcoming RRR region digital health barriers in Australia requires community engagement, device subsidies, digital literacy, and socioeconomic support for sustainable, accessible, and culturally appropriate care. Coordinated planning, cost management, and sustainable funding enhance efficiency, scalability, and long-term impact of digital health services in underserved RRR communities. Future research should co-design dietary interventions combining community-based support with digital approaches for greater effectiveness [50]. By aligning digital health initiatives with the realities of RRR region practice, Australia can move closer to bridging the health gap and ensuring equitable access to care for all communities.

Figure 3. The key implementation steps.



Acknowledgments

The authors sincerely thank the epidemiology team at the Rural Health Research Institute (RHRI), Charles Sturt University, which includes MMH, KYA, TS, AEA, MAM, MJAS, UKM, SBA, and SG, for their valuable insights and thoughtful suggestions.

Authors' Contributions

SM conceived the idea and conducted the literature review. SM extracted the data and wrote the first draft of the manuscript, which was critically revised by the Epi team and AR, focusing on improving cultural sensitivity, methodological appropriateness, respect for cultural values, and ensuring ethical considerations. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Rural and remote health. Australian Institute of Health and Welfare. 2024 Apr 30. URL: https://www.aihw.gov.au/reports/rural-remote-australians/rural-and-remote-health [accessed 2024-06-24]
2. Thapa S, Ahmed KY, Ross AG. Beyond statistics: health inequities in rural and remote communities of Australia. Lancet Public Health 2023 Nov;8(11):e834 [FREE Full text] [doi: 10.1016/S2468-2667(23)00246-3] [Medline: 37898513]
3. Australian Statistical Geography Standard (ASGS): volume 5 - remoteness structure, july 2016. Goverment of Australia. URL: https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/1270.0.55.005Main%20Features5July%202016 [accessed 2025-11-10]

XSL•FO
RenderX

4. Ion D. Cardihab enables home-based, digital cardiac rehab for remote North Queensland patients. Cardihab. 2025 Mar 21. URL: https://cardihab.com/news/mater-private-hospital-townsville-program-success/ [accessed 2025-10-28]
5. Australian burden of disease study 2022. Australian Institute of Health and Welfare. 2022 Dec 13. URL: https://www.aihw.gov.au/reports/burden-of-disease/australian-burden-of-disease-study-2022 [accessed 2024-06-06]
6. Digital health. Australian Institute of Health and Welfare. 2024 Jul 02. URL: https://www.aihw.gov.au/reports/australias-health/digital-health [accessed 2025-10-28]
7. Main contributors to the Indigenous life expectancy gap. Australian Institute of Health and Welfare. 2016. URL: https://www.aihw.gov.au/getmedia/5d39a104-a2d5-4ab5-900c-697ee0e5a1d8/ah16-5-8-main-contributors-indigenous-life-expectancy-gap.pdf.aspx [accessed 2025-10-28]
8. Health workforce. Australian Institute of Health and Welfare. 2024 Jul 02. URL: https://www.aihw.gov.au/reports/workforce/health-workforce [accessed 2025-10-28]
9. Admitted patients. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/hospitals/topics/admitted-patient-care [accessed 2025-10-28]
10. Butzner M, Cuffee Y. Telehealth interventions and outcomes across rural communities in the United States: narrative review. J Med Internet Res 2021 Aug 26;23(8):e29575 [FREE Full text] [doi: 10.2196/29575] [Medline: 34435965]
11. Leveraging digital technology in healthcare. Australian Government Productivity Commission. 2024 May 07. URL: https://www.pc.gov.au/inquiries-and-research/digital-healthcare/ [accessed 2025-10-28]
12. Whitacre BE. Rural EMR adoption rates overtake those in urban areas. J Am Med Inform Assoc 2015 Mar;22(2):399-408 [FREE Full text] [doi: 10.1093/jamia/ocu035] [Medline: 25665701]
13. Murphy P, Burge F, Wong ST. Measurement and rural primary health care: a scoping review. Rural Remote Health 2019 Aug;19(3):4911 [FREE Full text] [doi: 10.22605/RRH4911] [Medline: 31365828]
14. Nguyen KH, Wright C, Simpson D, Woods L, Comans T, Sullivan C. Economic evaluation and analyses of hospital-based electronic medical records (EMRs): a scoping review of international literature. NPJ Digit Med 2022 Mar 08;5(1):29 [FREE Full text] [doi: 10.1038/s41746-022-00565-1] [Medline: 35260765]
15. Virtual service expanding to relieve hospital pressures. Premier of Victoria. 2022 Apr 20. URL: https://www.premier.vic.gov.au/virtual-service-expanding-relieve-hospital-pressures [accessed 2025-10-28]
16. Australian-first virtual ED triage launches. Northern Health. 2020 Oct 08. URL: https://www.nh.org.au/australian-first-virtual-ed-triage-launches/ [accessed 2025-10-28]
17. Improving outcomes for people with cardiovascular disease. Cardihab. 2025. URL: https://cardihab.com/ [accessed 2025-10-28]
18. An overview of THIS WAY UP 2015 to 2025. Australian Government Productivity Commission. 2025. URL: https://assets.pc.gov.au/391878/sub036-mental-health-review-attachment.pdf?VersionId=aW_vZfrkuuvukzQfXUOq9I5_Ny5tD4Tq [accessed 2025-10-28]
19. Understanding blended care. Australian Government Department of Health, Disability and Ageing. 2022. URL: https://www.medicarementalhealth.gov.au/for-health-professionals/understanding-blended-care [accessed 2025-10-28]
20. Australia's first virtual hospital. Sydney Local Health District. 2025. URL: https://slhd.health.nsw.gov.au/sydney-virtual/about/history#sydney-local-health-district-virtual-hospital,-an-australian-first [accessed 2025-10-28]
21. Virtual clinical care. South Australia Health. 2025. URL: https://www.sahealth.sa.gov.au/wps/wcm/connect/public+content/sa+health+internet [accessed 2025-10-28]
22. AI trends for healthcare. Commonwealth Scientific and Industrial Research Organisation. 2024 Mar. URL: https://aehrc.csiro.au/wp-content/uploads/2024/03/AI-Trends-for-Healthcare.pdf [accessed 2025-10-28]
23. Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. NPJ Digit Med 2019 Aug 20;2:79 [FREE Full text] [doi: 10.1038/s41746-019-0158-1] [Medline: 31453374]
24. Maslej N, Fattorini L, Brynjolfsson E, Etchemendy J, Ligett K, Lyons T, et al. Artificial intelligence index report 2023. arXiv. Preprint posted online on October 05, 2023 2025 [FREE Full text] [doi: 10.48550/arXiv.2310.03715]
25. Chatterjee J, Sujatha R. Transforming healthcare: the synergy of telemedicine, telehealth, artificial intelligence. In: Chatterjee JM, Sujatha R, Saxena SK, editors. Role of Artificial Intelligence, Telehealth, and Telemedicine in Medical Virology. Singapore, Singapore: Springer; Feb 01, 2025:1-29.
26. Digital health and connectivity in rural Australia. National Rural Health Alliance. 2024. URL: https://www.ruralhealth.org.au/publication/fact-sheet/digital-health-and-connectivity-in-rural-australia/ [accessed 2025-10-28]
27. Henry TA. 7 EHR usability, safety challenges—and how to overcome them. American Medical Association. 2023 Dec 11. URL: http://ama-assn.org/practice-management/digital-health/7-ehr-usability-safety-challenges-and-how-overcome-them [accessed 2025-10-28]
28. Lorenzi NM, Kouroubali A, Detmer DE, Bloomrosen M. How to successfully select and implement electronic health records (EHR) in small ambulatory practice settings. BMC Med Inform Decis Mak 2009 Feb 23;9:15 [FREE Full text] [doi: 10.1186/1472-6947-9-15] [Medline: 19236705]
29. Holt M, MacGibbon J, Smith AK, Broady TR, Davis MD, Newman CE. Knowledge of Australia's My Health Record and factors associated with opting out: results from a national survey of the Australian general population and communities

XSL•FO
RenderX

affected by HIV and sexually transmissible infections. PLOS Digit Health 2023 Mar 01;2(3):e0000200 [FREE Full text] [doi: 10.1371/journal.pdig.0000200] [Medline: 36857326]
30. de Mesquita RC, Edwards I. Systematic literature review of my health record system. Asia Pac J Health Manag 2020 Mar 15;15(1):14-25 [FREE Full text] [doi: 10.24083/apjhm.v15i1.311]
31. Connecting Australia to a healthier future. DXC Technology. URL: https://dxc.com/au/en/insights/customer-stories/connecting-australia-to-a-healthier-future [accessed 2025-10-28]
32. What is healthdirect video call? Healthdirect Australia. URL: https://help.vcc.healthdirect.org.au/about-healthdirect-video-call/whatishealthdirectvideocall [accessed 2025-10-28]
33. Free, digital psychological care with MindSpot. MindSpot. URL: https://www.mindspot.org.au/ [accessed 2025-10-28]
34. Bourke SL, Harper C, Johnson E, Green J, Anish L, Muduwa M, et al. Health care experiences in rural, remote, and metropolitan areas of Australia. Online J Rural Nurs Health Care 2021 May 04;21(1):67-84 [FREE Full text] [doi: 10.14574/ojrnhc.v21i1.652]
35. McMahon B. Healthdirect – taking the difficulty out of navigating health services. Health Voices. 2022. URL: https://healthvoices.org.au/issues/may-2022/healthdirect-taking-the-difficulty-out-of-navigating-health-services/ [accessed 2025-10-28]
36. Breadon P, Geraghty J, Jones D, Baldwin E. Special treatment: improving Australians' access to specialist care. Grattan Institute. 2025 Jun 15. URL: https://grattan.edu.au/report/special-treatment-improving-australians-access-to-specialist-care/ [accessed 2025-10-28]
37. Healthdirect Australia 2021-24 strategic plan. Royal Australian College of General Practitioners. 2021 May 28. URL: https://www.racgp.org.au/FSDEDEV/media/documents/RACGP/Reports%20and%20submissions/2021/Healthdirect-2021-24-Strategic-Plan-consultation.pdf [accessed 2025-10-28]
38. Ross J, Stevenson F, Dack C, Pal K, May C, Michie S, et al. Developing an implementation strategy for a digital health intervention: an example in routine healthcare. BMC Health Serv Res 2018 Oct 19;18(1):794 [FREE Full text] [doi: 10.1186/s12913-018-3615-7] [Medline: 30340639]
39. Field P, Franklin RC, Barker R, Ring I, Leggat PA. Cardiac rehabilitation in rural and remote areas of North Queensland: how well are we doing? Aust J Rural Health 2022 Aug;30(4):488-500 [FREE Full text] [doi: 10.1111/ajr.12861] [Medline: 35298054]
40. Gentili A, Failla G, Melnyk A, Puleo V, Tanna GL, Ricciardi W, et al. The cost-effectiveness of digital health interventions: a systematic review of the literature. Front Public Health 2022 Aug 11;10:787135 [FREE Full text] [doi: 10.3389/fpubh.2022.787135] [Medline: 36033812]
41. Digital health blueprint. Department of Health and Aged Care, Commonwealth of Australia. URL: https://www.health.gov.au/sites/default/files/2024-01/the-digital-health-blueprint-and-action-plan-2023-2033_0.pdf [accessed 2025-10-28]
42. Heart, stroke and vascular disease: Australian facts. Australian Institute of Health and Welfare. 2025 Oct 22. URL: https://www.aihw.gov.au/reports/heart-stroke-vascular-diseases/hsvd-facts/contents/about [accessed 2025-10-28]
43. 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 01;68(5):427-429 [FREE Full text] [doi: 10.1176/appi.ps.201600541] [Medline: 28412890]
44. Titov N, Dear BF, Staples LG, Bennett-Levy J, Klein B, Rapee RM, et al. MindSpot clinic: an accessible, efficient, and effective online treatment service for anxiety and depression. Psychiatr Serv 2015 Oct;66(10):1043-1050. [doi: 10.1176/appi.ps.201400477] [Medline: 26130001]
45. Battersby MW, Baigent MF, Redpath P. MindSpot: a valuable service that raises questions. Lancet Digit Health 2020 Nov;2(11):e562-e563 [FREE Full text] [doi: 10.1016/S2589-7500(20)30248-X] [Medline: 33328083]
46. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-mental health care in Australia: a systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]
47. Prescott MR, Sagui-Henson SJ, Welcome Chamberlain CE, Castro Sweet CC, Altman M. Real world effectiveness of digital mental health services during the COVID-19 pandemic. PLoS One 2022 Aug 18;17(8):e0272162 [FREE Full text] [doi: 10.1371/journal.pone.0272162] [Medline: 35980879]
48. Rural areas desperate for more local mental health services. University of Western Australia. URL: https://www.uwa.edu.au/news/article/2019/august/rural-areas-desperate-for-more-local-mental-health-services [accessed 2025-10-28]
49. Type 2 diabetes in young Aboriginal people clinical. Northern Territory Health. URL: https://health.nt.gov.au/media/pdf/chronic-conditions/type-2-diabetes-in-young-aboriginal-people-clinical-policy.pdf [accessed 2025-10-28]
50. Needham C, Partridge SR, Alston L, Rawstorn JC, Livingstone KM. Co-designing interventions to improve diets in rural communities. Proc Nutr Soc 2025 Feb 06:1-7. [doi: 10.1017/S0029665125000060] [Medline: 39910988]
51. Shaddock L, Smith T. Potential for use of portable ultrasound devices in rural and remote settings in Australia and other developed countries: a systematic review. J Multidiscip Healthc 2022 Mar 29;15:605-625 [FREE Full text] [doi: 10.2147/JMDH.S359084] [Medline: 35378744]

XSL•FO
RenderX

52. Blakely B, Magrabi MF. Healthdirect Triage Platform usability follow up study: second usability study report. Macquarie University. 2025. URL: https://researchers.mq.edu.au/en/publications/healthdirect-triage-platform-usability-follow-up-study-second-usa/ [accessed 2025-10-28]
53. About my health record. Healthdirect Australia. 2024. URL: https://www.healthdirect.gov.au/my-health-record [accessed 2025-10-28]
54. Action plan for the digital health blueprint 2023-2033. Australian Government Department of Health and Aged Care. 2023. URL: https://www.health.gov.au/sites/default/files/2023-12/the_action_plan_for_the_digital_health_blueprint_2023-2033.pdf [accessed 2025-10-28]
55. Independent evaluation of supported digital mental health services – phase 2 final report. Australian Government Department of Health, Disability, and Aging. 2022. URL: https://www.health.gov.au/resources/publications/independent-evaluation-of-supported-digital-mental-health-services-phase-2-final-report?language=en [accessed 2025-10-28]
56. Get health information and advice. Healthdirect Australia. URL: https://www.healthdirect.gov.au/ [accessed 2025-10-28]
57. About us. Cardihab. 2024. URL: https://cardihab.com/who-we-are/ [accessed 2025-10-28]
58. Our evidence. Cardihab. URL: https://cardihab.com/our-evidence/ [accessed 2025-10-28]
59. Meeting Australians where they are. MindSpot. URL: https://www.mindspot.org.au/rural-mental-health/ [accessed 2025-10-28]
60. MindSpot clinic. Healthdirect Australia. URL: https://www.healthdirect.gov.au/partners/mindspot-clinic [accessed 2025-10-28]
61. eMHPrac: supporting national implementation of digital mental health in Australia. eMental Health International Collaborative. 2025 May 20. URL: https://emhicglobal.com/case-studies/emhprac-supporting-national-implementation-of-digital-mental-health-in-australia/ [accessed 2025-10-28]
62. OMA's virtual care working group FAQ. The Ontario Medical Association. 2023. URL: https://www.oma.org/siteassets/oma/media/pagetree/pps/running/operations/virtual-care/vc-working-group-faq.pdf [accessed 2025-10-28]
63. Virtual care policy framework. Government of Canada. 2021 Jul 07. URL: https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/bilateral-agreement-pan-canadian-virtual-care-priorities-covid-19/policy-framework.html [accessed 2025-10-28]
64. Ontario's virtual care streamlined action plan. Government of Canada. 2022. URL: https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/bilateral-agreement-pan-canadian-virtual-care-priorities-covid-19/action-plan-ontario.html [accessed 2025-10-28]
65. OuluHealth facilitates close cooperation that drives innovation, boosts the healthtech business and supports the implementation of better healthcare services. OULUHEALTH. URL: https://ouluhealth.fi/about/stakeholders/ [accessed 2025-11-10]
66. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
67. Statistical information and data. Services Australia. URL: https://www.servicesaustralia.gov.au/statistical-information-and-data?context=22 [accessed 2025-10-28]
68. Vayalil MP. 6 key metrics for HCPs in value-based care. blueBriX. 2025 Jan 08. URL: https://bluebrix.health/blogs/6-key-metrics-for-hcps-in-value-based-care [accessed 2025-10-28]
69. Ravi S, Meyerowitz-Katz G, Yung C, Ayre J, McCaffery K, Maberly G, et al. Effect of virtual care in type 2 diabetes management - a systematic umbrella review of systematic reviews and meta-analysis. BMC Health Serv Res 2025 Mar 06;25(1):348 [FREE Full text] [doi: 10.1186/s12913-025-12496-0] [Medline: 40050767]
70. Clough B, Fraser J, Flemington T, Power T. Implementation and evaluation of cultural safety initiatives in Australian hospital settings: a scoping review. J Transcult Nurs 2025 May;36(3):279-289 [FREE Full text] [doi: 10.1177/10436596241296818] [Medline: 39559851]
71. Your my health record. Australian Government. 2024. URL: https://my.gov.au/en/services/health-and-disability/seeking-medical-help/accessing-your-medical-information/your-my-health-record [accessed 2025-10-28]
72. About MindSpot. MindSpot. 2025. URL: https://www.mindspot.org.au/about-mindspot/ [accessed 2025-10-28]
73. Rural community engagement resource guide. Rural Health Value. 2023. URL: https://ruralhealthvalue.public-health.uiowa.edu/files/RHV_Community%20Engagement%20Resource%20Guide.pdf [accessed 2025-10-28]
74. Osman S, Churruca K, Ellis LA, Braithwaite J. Beyond the planned and expected: the unintended consequences of telehealth in rural and remote Australia through a complexity lens. Med J Aust 2024 Jun 03;220(10):496-498 [FREE Full text] [doi: 10.5694/mja2.52294] [Medline: 38703008]
75. Hamilton S, Mills B, McRae S, Thompson S. Evidence to service gap: cardiac rehabilitation and secondary prevention in rural and remote Western Australia. BMC Health Serv Res 2018 Jan 30;18(1):64 [FREE Full text] [doi: 10.1186/s12913-018-2873-8] [Medline: 29382343]
76. Rural health in Australia SNAPSHOT 2023. National Rural Health Alliance. 2023. URL: https://www.ruralhealth.org.au/wp-content/uploads/2024/04/NRHA_rural_health_in_Australia_snapshot_2023.pdf [accessed 2025-10-28]
77. Krahe MA, Baker S, Woods L, Larkins SL. Factors that influence digital health implementation in rural, regional, and remote Australia: an overview of reviews and recommended strategies. Aust J Rural Health 2025 Apr;33(2):e70045. [doi: 10.1111/ajr.70045] [Medline: 40202368]

XSL•FO
RenderX

78. Digital exclusion in Australia. Australian Government Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts. 2025 Aug. URL: https://www.infrastructure.gov.au/sites/default/files/documents/digital-exclusion-in-australia-report.pdf [accessed 2025-10-28]
79. Digital health literacy key to overcoming barriers for health workers, WHO study says. World Health Organization. 2023 Sep 18. URL: https://www.who.int/europe/news/item/18-09-2023-digital-health-literacy-key-to-overcoming-barriers-for-health-workers--who-study-says [accessed 2025-10-28]
80. Closing the gap in a generation: health equity through action on the social determinants of health. World Health Organization. 2008 Aug 27. URL: https://iris.who.int/server/api/core/bitstreams/cb08095c-55c8-484e-bff6-0e9c78fd38dd/content [accessed 2025-10-28]
81. Home Care Packages: access to $2,500 for urgent goods, equipment, and assistive technology. Australian Government Department of Health, Disability and Ageing. URL: https://www.health.gov.au/resources/publications/home-care-packages-access-to-2500-for-urgent-goods-equipment-and-assistive-technology?language=en [accessed 2025-10-28]
82. Specialist training program. Australian Government Department of Health, Disability and Ageing. URL: https://www.health.gov.au/our-work/specialist-training-program [accessed 2025-10-28]

Abbreviations

EHR: electronic health record
EMR: electronic medical record
GP: general practitioner
RRR: rural, regional, and remote

Edited by N Cahill; submitted 11.10.24; peer-reviewed by M Adamakis, E Popa, A Hasnain; comments to author 09.04.25; revised version received 29.08.25; accepted 11.09.25; published 02.12.25.

Please cite as:
Mahmood S, Huda MM, Yimam Ahmed K, Subash T, Astawesegn FH, Anyasodor AE, Moni MA, J A Shiddiky M, K Mondal U, Aychiluhm SB, Giri S, Ross A
Digital Health for Australia: Bridging the Rural, Regional, and Remote Health Gap
Interact J Med Res 2025;14:e67460
URL: https://www.i-jmr.org/2025/1/e67460
doi:10.2196/67460
PMID:

©Shakeel Mahmood, M Mamun Huda, Kedir Yimam Ahmed, Thapa Subash, Feleke Hailemichael Astawesegn, Anayochukwu Edward Anyasodor, Mohammad Ali Moni, Muhammad J A Shiddiky, Utpal K Mondal, Setognal Birara Aychiluhm, Santosh Giri, Allen Ross. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 02.12.2025. 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

Acupuncture and Moxibustion for Poststroke Depression: Systematic Review

Lu Meng[1], MMed; Chuang-Long Xu[1], MMed; Xiao-Xu He[2], MMed; Xiao-Chan Tan[3], MD, PhD

[1]Ningxia Hui Autonomous Region Hospital of Traditional Chinese Medicine, Ningxia Hui Autonomous Region Academy of Traditional Chinese Medicine, Yinchuan, China

[2]Hangzhou Red Cross Hospital (Zhejiang Hospital of Integrated Traditional and Western Medicine), Hangzhou, China

[3]The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), No. 54, Post Road, Shangcheng District, Zhejiang, Hangzhou, China

Corresponding Author:
Xiao-Chan Tan, MD, PhD
The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), No. 54, Post Road, Shangcheng District, Zhejiang, Hangzhou, China

Abstract

Background: Poststroke depression (PSD) is a common complication following stroke. In recent years, several systematic reviews have evaluated the effects of moxibustion and acupuncture on PSD; however, their findings have been inconsistent.

Objective: This overview of systematic reviews aimed to assess the methodological quality, reporting quality, and strength of evidence of existing systematic reviews on acupuncture and moxibustion for PSD. In addition, this study also analyzed the limitations of previous studies and suggested directions for future research.

Methods: Systematic reviews concerning acupuncture and moxibustion for PSD published before August 10, 2024, were identified from 8 databases, including PubMed, Embase, Cochrane Library, Web of Science, China National Knowledge Infrastructure, Wanfang, VIP Database, and Chinese Biomedical Literature Database. Eligible studies included systematic reviews and meta-analyses of randomized controlled trials comparing moxibustion and acupuncture for the treatment of PSD. The methodological quality, reporting quality, and evidence quality were evaluated using AMSTAR 2 (Assessment of Multiple Systematic Reviews-2), PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020, and GRADE (Grading of Recommendations Assessment, Development and Evaluation), respectively. The corrected covered area was calculated to assess overlap among the included primary studies.

Results: A total of 24 studies were included. According to the Assessment of Multiple Systematic Reviews-2 assessment, all studies were rated as having “low” or “critically low” methodological quality. Based on PRISMA, 1 study demonstrated seriously inadequate reporting quality, while 21 studies had partially inadequate reporting quality. The quality of evidence in the included reviews ranged from very low to moderate. Most of the primary outcomes exhibited mild to moderate overlap among studies.

Conclusions: Most of the included systematic reviews indicated that acupuncture may be beneficial for PSD. Nevertheless, the methodology, reporting, and evidence quality of these reviews require improvement. Stronger evidence will depend on the conduct of larger, multicenter, rigorously designed randomized controlled trials, as well as high-quality systematic reviews.

Trial Registration: PROSPERO CRD42024576753; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024576753

(Interact J Med Res 2025;14:e76577) doi:10.2196/76577

KEYWORDS

acupuncture; humans; poststroke depression; systematic review; moxibustion

Introduction

Poststroke depression (PSD) is a prevalent neuropsychiatric complication following stroke, characterized by depressed mood, loss of interest, and diminished quality of life during the recovery process [1]. PSD not only impairs patients’ quality of life but also exacerbates cognitive dysfunction, delays neurological recovery, and elevates the risk of recurrent stroke and mortality [2]. The pathogenesis of PSD remains incompletely understood but is thought to involve a combination of stroke-induced damage to emotional regulatory brain regions, resulting in neurotransmitter imbalances, reduced ability to perform daily functions, and insufficient social support [3].

Globally, the prevalence of PSD ranges from 17% to 34.3% [4]. Its incidence is influenced by factors such as age, gender, education level, marital and economic status, and stroke type [5]. The mainstays of Western medical management for PSD include pharmacotherapy, most notably selective serotonin

XSL•FO
RenderX

reuptake inhibitors and norepinephrine-dopamine reuptake inhibitors, and psychotherapeutic approaches such as cognitive-behavioral and interpersonal therapies [6,7]. However, pharmacotherapy is frequently associated with adverse effects such as dry mouth, constipation, dizziness, somnolence, sexual dysfunction, and, in some patients, lack of efficacy or development of drug resistance [8-10].

Acupuncture, an integral part of traditional Chinese medicine, has demonstrated beneficial effects for PSD. Clinical evidence suggests that acupuncture can alleviate depressive symptoms, improve quality of life, and reduce the adverse effects induced by pharmacological agents [11,12]. The putative mechanisms include modulation of neurotransmitter systems, enhancement of cerebral blood flow, and regulation of immune function [13]. Numerous clinical trials and case studies point to acupuncture and moxibustion as effective interventions for PSD [14,15]. In recent years, multiple systematic reviews have synthesized the evidence on acupuncture and moxibustion for PSD, occasionally yielding inconsistent or conflicting results [16,17].

Therefore, this study presents an overview of systematic reviews that comprehensively summarizes the current state of research on acupuncture and moxibustion for PSD. By synthesizing existing evidence, it aims to provide a scientific foundation for clinical practice, clarify methodological limitations of previous studies, and propose recommendations for future research.

Methods

Protocol and Registration

The protocol for this overview of systematic reviews was registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42024576753). The reporting of this study adheres to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 guidelines [18].

Data Sources and Search Strategy

A total of 8 electronic databases were searched: PubMed, Embase, Cochrane Library, Web of Science, China National Knowledge Infrastructure, Wanfang, VIP Database, and Chinese Biomedical Literature Database. The search included records published up to August 10, 2024. Keywords used were acupuncture, moxibustion, electroacupuncture, manual acupuncture, acupoint catgut embedding, systematic review, meta-analysis, and PSD. Detailed search strategies for each database are provided in Table S1 in Multimedia Appendix 1.

Eligibility Criteria

Studies were included if they met the following criteria: (1) systematic reviews or meta-analyses of randomized controlled trials (RCTs) or quasi-RCTs focused on acupuncture and moxibustion for PSD; (2) participants diagnosed with PSD, confirmed by computed tomography or magnetic resonance imaging, regardless of gender, age, ethnicity, education, or economic status; (3) interventions applied to the experimental group included acupuncture, moxibustion, electroacupuncture, manual acupuncture, acupoint catgut embedding, or transcutaneous electrical acupoint stimulation, alone or in combination with Western medicine or basic medications. Control groups received interventions other than acupuncture or moxibustion (eg, Western medicine, placebo, and sham acupuncture) to compare the efficacy of acupuncture with other treatment methods or placebo in the treatment of PSD; (4) main outcome indicators included overall effectiveness rate, Hamilton Depression Rating Scale (HAMD), Self-Rating Depression Scale (SDS), and the modified Edinburgh-Scandinavian stroke scale scores; and (5) no restrictions were imposed on the duration of the follow-up period (Table S2 in Multimedia Appendix 1 [16,19-41]).

Exclusion criteria were (1) systematic reviews or meta-analyses that did not include RCTs; (2) studies involving patients without a clear diagnosis of PSD; (3) control group interventions that included acupuncture; (4) duplicate publications; (5) protocols for systematic reviews, commentaries, or conference abstracts; and (6) studies lacking complete or essential information.

Literature Selection

All studies identified in the search were imported into EndNote (EndNote X20; Clarivate) for deduplication. Two independent reviewers (LM and CLX) screened titles and abstracts according to the inclusion and exclusion criteria and further independently assessed the full text of potentially eligible studies. Disagreements were resolved by discussion, with a third reviewer (XCT) involved when necessary.

Data Extraction

Data extraction was conducted independently by 2 reviewers (LM and CLX). Extracted data included (1) bibliographic details (title, authors, and year of publication); (2) number and sample size of included studies; (3) intervention methods, control interventions, outcome results, 95% CI, I^2 statistics, P values, and study conclusions; and (4) methodological quality assessments of the included studies. Any discrepancies were resolved by discussion with a third reviewer (XCT).

Quality Assessment

Methodological Quality

The included systematic reviews were assessed using the Assessment of Multiple Systematic Reviews-2 (AMSTAR 2) tool [42]. Items 2, 4, 7, 9, 11, 13, and 15 were considered critical domains. Reviews were categorized as “high” quality (no noncompliance on critical items), “moderate” quality (noncompliance on noncritical items only), “low” quality (1 critical item noncompliant), and “very low” quality (more than 1 critical item noncompliant).

In addition, we used the Cochrane Risk-of-Bias tool to reassess the risk of bias in the original RCTs included in the systematic reviews and performed statistical analyses on items with a high risk of bias, such as random sequence generation, allocation concealment, and blinding.

Reporting Quality

PRISMA 2020 [18] was used to evaluate reporting quality. Each of the 27 items was scored as 1 (fully reported), 0.5 (partially reported), or 0 (not reported), for a maximum score of 27. Scores <15 indicated “severely deficient” reporting, 15.5 - 21 “somewhat deficient,” and 21.5 - 27 “relatively complete” [43].

XSL•FO
RenderX

Evidence Quality

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) system [44,45] was used to assess the quality of each outcome. Evidence was categorized as high, moderate, low, or very low, in accordance with Cochrane and contemporary guidelines [46]. Quality assessments were carried out independently by 2 reviewers (LM and XXH), with disagreements resolved through discussion or consultation with a third reviewer (XCT).

Heterogeneity Assessment

Statistical heterogeneity among studies was assessed using the I^2 statistic, quantifying the percentage of variability due to heterogeneity rather than chance [47,48]. I^2 values of 25%, 50%, and 75% were considered low, moderate, and high heterogeneity, respectively [49].

Data Synthesis and Analysis

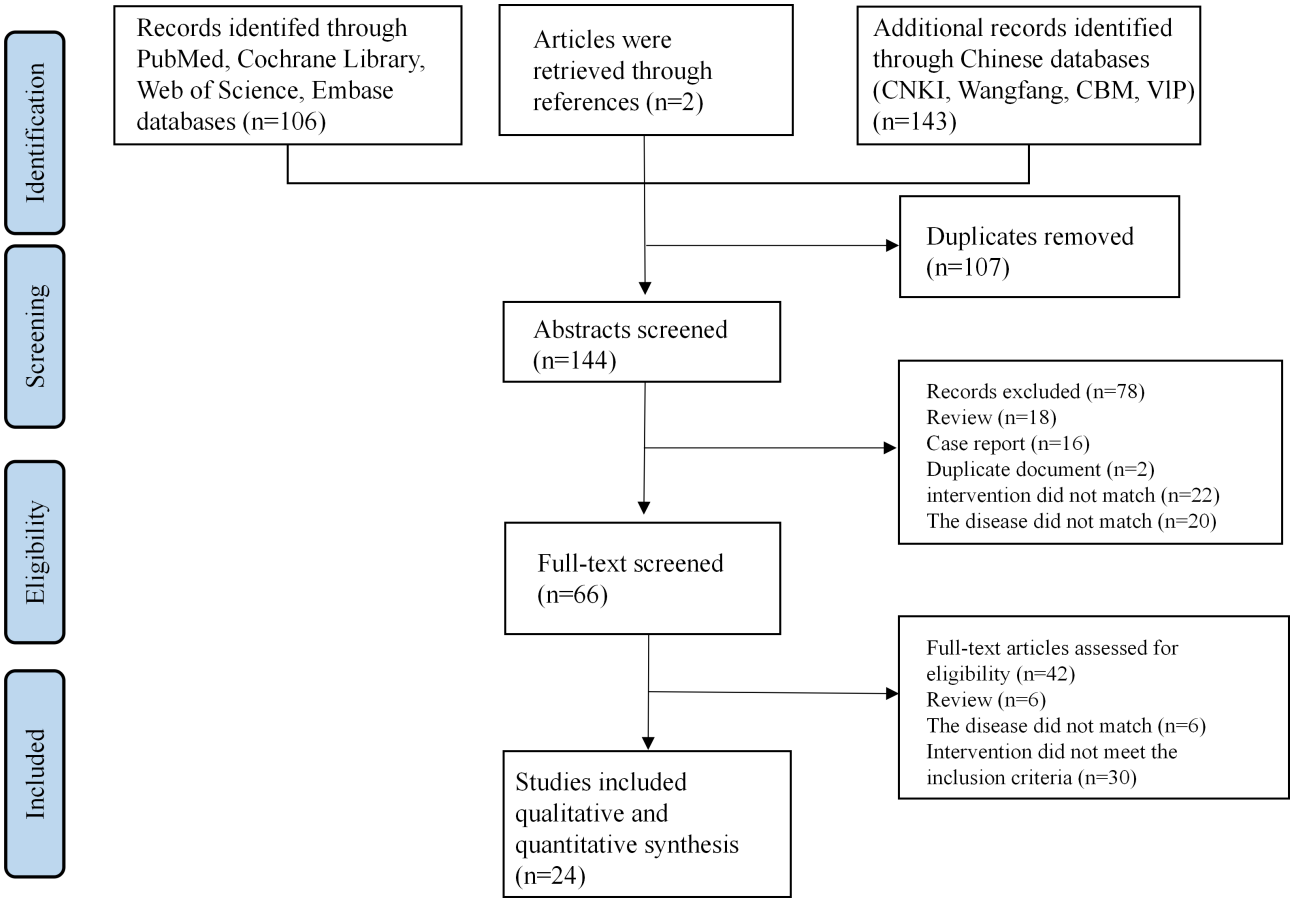
A narrative synthesis was performed, focusing on study population, sample size, interventions, outcome indicators, relative effect sizes, I^2 statistics, *P* values, and main conclusions. Where possible, results were also summarized in a tabular form. All collected data were entered into Microsoft Excel 2019 for qualitative synthesis. Overlap of primary studies was assessed using the corrected covered area (CCA) method and calculated as follows: CCA (%)=(N-r)/(rc-r), where N is the total number of included observations from all reviews, r is the number of unique primary studies, and c is the number of systematic reviews. Interpretation was categorized as 0% - 5% (slight overlap), 6% - 10% (moderate overlap), 11% - 15% (high overlap), and >15% (very high overlap).

Results

Literature Search Results

A total of 251 records were screened for eligibility. After removing 107 duplicates, 144 articles remained. Upon review of titles and abstracts, 78 articles were excluded, with an additional 42 excluded after full-text assessment. Ultimately, 24 articles [16,19-41] met the inclusion criteria for this overview (Figure 1).

Figure 1. Flow diagram of the search and screening of the studies included in this overview. CBM: Chinese Biomedical Literature Database; CNKI: China National Knowledge Infrastructure.



Study Characteristics

The included studies comprised 24 systematic reviews and meta-analyses, with 11 papers published in English and the remaining in Chinese [20,21,24-26,28,29,31,33,37-39,41] (Table 1). Interventions in the experimental group included acupuncture, electroacupuncture, scalp acupuncture, alone or in combination with Western medicine. Comparator groups received Western medicine, sham acupuncture, placebo, or no treatment. All systematic reviews evaluated the methodological quality of their included studies, with 17 [16,20,23-26,28-30,32,34,35,38-41] using the Cochrane Collaboration Risk of Bias tool, 3 [19,21,33] using the Jadad scale, and 3 [27,31,37] combining both tools. A total of 21

XSL•FO
RenderX

systematic reviews [16,19,21,23-26,28-41] concluded that acupuncture was effective for the treatment of PSD, while 3 [20,22,27] reviews found insufficient evidence to support its benefits. Acupuncture was associated with fewer adverse events than antidepressants.

XSL•FO
RenderX

Table . Baseline characteristics of included reviews.

Study	Trials (subjects)	Experimental intervention	Control intervention	Result summary	Quality assessment
Zhang et al (2010) [19]	14 (1512)	AT[a]	WM[b]/WC[c]	Acupuncture therapy is safe and effective in treating PSD[d], and could be considered an alternative option for the disorder.	Jadad
Xiong et al (2010) [20]	20 (2031)	AT/EA[e]+WM	WM	Acupuncture is not inferior to Western medicine, and it is worth noting that acupuncture is associated with few adverse reactions.	Cochrane
Lai et al (2012) [21]	29 (2394)	AT/EA+WM	WM	Electroacupuncture combined with fluoxetine improves depression status more significantly than fluoxetine alone in patients with PSD.	Jadad
Zhang et al (2012) [22]	15 (1096)	AT	WM	Comparison between the acupuncture group and Western medicine group in treating PSD revealed that there is a statistical difference in curative rate and remarkably effective rate, but no difference in effective rate.	Grade scale
Zhang et al (2014) [23]	13 (845)	AT/EA/MT[f]+WM	WM	Treating PSD with acupuncture was more effective compared with Western medicine, but the result was less reliable and the quality of evidence was poor.	Cochrane
Tan et al (2016) [24]	14 (1180)	EA	WM	EA may be a more effective and safer treatment for PSD than antidepressants.	Cochrane
Wang et al (2016) [25]	27 (1729)	AT/AT+WM	WM	The therapeutic effect of acupuncture in the treatment of PSD is superior to that of Western medicine	Cochrane
Li et al (2017) [26]	10 (806)	AT+WM	WM	Acupuncture treatment of PSD is more efficient than fluoxetine hydrochloride treatment.	Cochrane
Li et al (2018) [27]	18 (1536)	EA	WM/SAT[g]/UC[h]	There was no significant difference between EA and antidepressants in the severity of depression; however, EA caused fewer adverse events than antidepressants.	Cochrane+ Jadad

XSL•FO
RenderX

Study	Trials (subjects)	Experimental intervention	Control intervention	Result summary	Quality assessment
Huang et al (2018) [28]	13 (1193)	AT/EA+WM	WM	Acupuncture and electro-acupuncture are effective therapy methods in improving the state of depression of patients with PSD, and they can obviously improve the effect of treatment when combined with Western medicine.	Cochrane
Que et al (2018) [29]	18 (1813)	AT/EA+WM	WM	Compared with Prozac, acupuncture treatment for PSD had fewer adverse effects and fewer symptoms, but there was little difference in clinical efficacy and reduction in HAMD[i] scores.	Cochrane
Zhang et al (2019) [30]	7 (514)	AT/MT	WM/CHM[j]	Acupuncture is an effective and safe treatment for PSD.	Cochrane
Xu et al (2019) [31]	19 (1542)	AT+WM	SAT/MT	Acupuncture is an effective and safe treatment for PSD. The acupuncture group is more effective than the conventional drug group.	Cochrane+ Jadad
Zhang et al (2020) [32]	13 (904)	AT+WM/EA+WM	WM	Acupuncture combined with antidepressants is an effective treatment for PSD.	Cochrane
Yin et al (2020) [33]	15 (1503)	AT	WM	Acupuncture is more effective than conventional medication in treating PSD.	Jadad
Liu et al (2021) [34]	17 (1402)	AT/EA	WM/SAT/UC	Acupuncture could reduce the degree of PSD.	Cochrane
Zhang et al (2021) [35]	14 (1124)	AT/EA	WM	Acupuncture not only can reduce the severity of PSD, but also has significant effects on decreasing the appearance of other adverse events	Cochrane
Wang et al (2021) [36]	19 (1606)	EA	WM	Compared with antidepressants, electroacupuncture is not less effective at improving depression symptoms in patients with PSD, with greater safety.	Cochrane

XSL•FO
RenderX

Study	Trials (subjects)	Experimental intervention	Control intervention	Result summary	Quality assessment
Wang et al (2021) [37]	8 (769)	AT/EA(+WM)	WM	Compared with fluoxetine, acupuncture has high clinical efficacy in treating PSD and has significant advantages in improving patients' symptoms and reducing adverse effects.	Cochrane+ Jadad
Zhang et al (2021) [38]	14 (1120)	AT/MT/AT+MT	WM	Acupuncture and moxibustion treatment are superior to antidepressants in improving the depressive state and activities of daily living of patients with PSD.	Cochrane
Lin et al (2022) [39]	16 (1618)	AT/AT+WM	WM	The Xing nao Kai qiao acupuncture therapy can significantly improve emotional well-being and neurological function in patients with PSD.	Cochrane
Zhong et al (2023) [16] []	11 (1225)	SA+CT[k]	CT	SA[l] combined with CT can effectively improve the treatment's success rate for PSD and reduce the severity of depressive symptoms measured by the Self-Rating Depression Scale.	Cochrane
Jiang et al (2023) [40]	14 (1263)	SA/EA+WM	WM	Scalp acupuncture has superior efficacy and safety compared to Western medicine for PSD.	Cochrane
Yang et al (2024) [41]	15 (1016)	EA/EA+WM	WM	Electroacupuncture for PSD can reduce HAMD, NIHSS[m], and SDS[n] scores, with significant clinical efficacy and no toxic side effects.	Cochrane

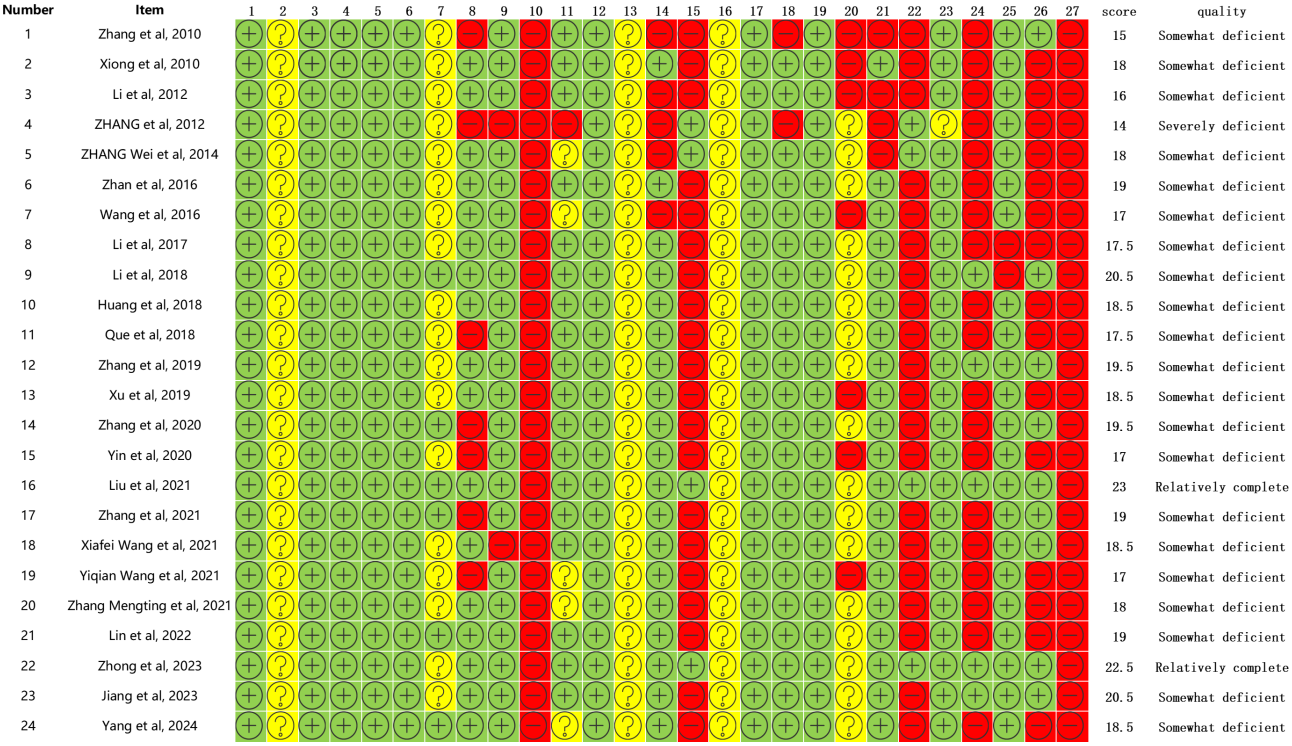
[a]AT: acupuncture therapy.
[b]WM: Western medicine.
[c]WC: waist circumference.
[d]PSD: poststroke depression.
[e]EA: electroacupuncture.
[f]MT: mirror therapy.
[g]SAT: sham acupuncture therapy.
[h]UC: ulcerative colitis.
[i]HAMD: Hamilton Depression Rating Scale.
[j]CHM: Chinese herbal medicine.
[k]CT: computed tomography.
[l]SA: scalp acupuncture.
[m]NIHSS: National Institutes of Health Stroke Scale.
[n]SDS: Self-Rating Depression Scale.

XSL•FO
RenderX

Methodological Appraisal

The methodological quality of the included reviews was assessed using the AMSTAR 2 tool. Every review had at least 1 critical domain rated as "very low" in methodological quality. In particular, major deficiencies were noted in protocol registration (Item 2), provision of excluded studies (Item 7), and comprehensive assessment of publication bias (Item 15; Figure 2). Out of the 24 included studies, 22 [16,19-26,28-39,41] were rated as "critically low" and 2 [27,40] as "low" in methodological quality. Notably, most studies (21/24) reporting positive findings for acupuncture had critically low methodological quality, while all 3 reviews [20,27,29] indicating insufficient evidence also had methodological shortcomings.

Figure 2. Results of the methodological quality evaluation using Assessment of Multiple Systematic Reviews-2 (AMSTAR 2). "+" denotes yes, "–" denotes no, and "?" denotes partial yes.

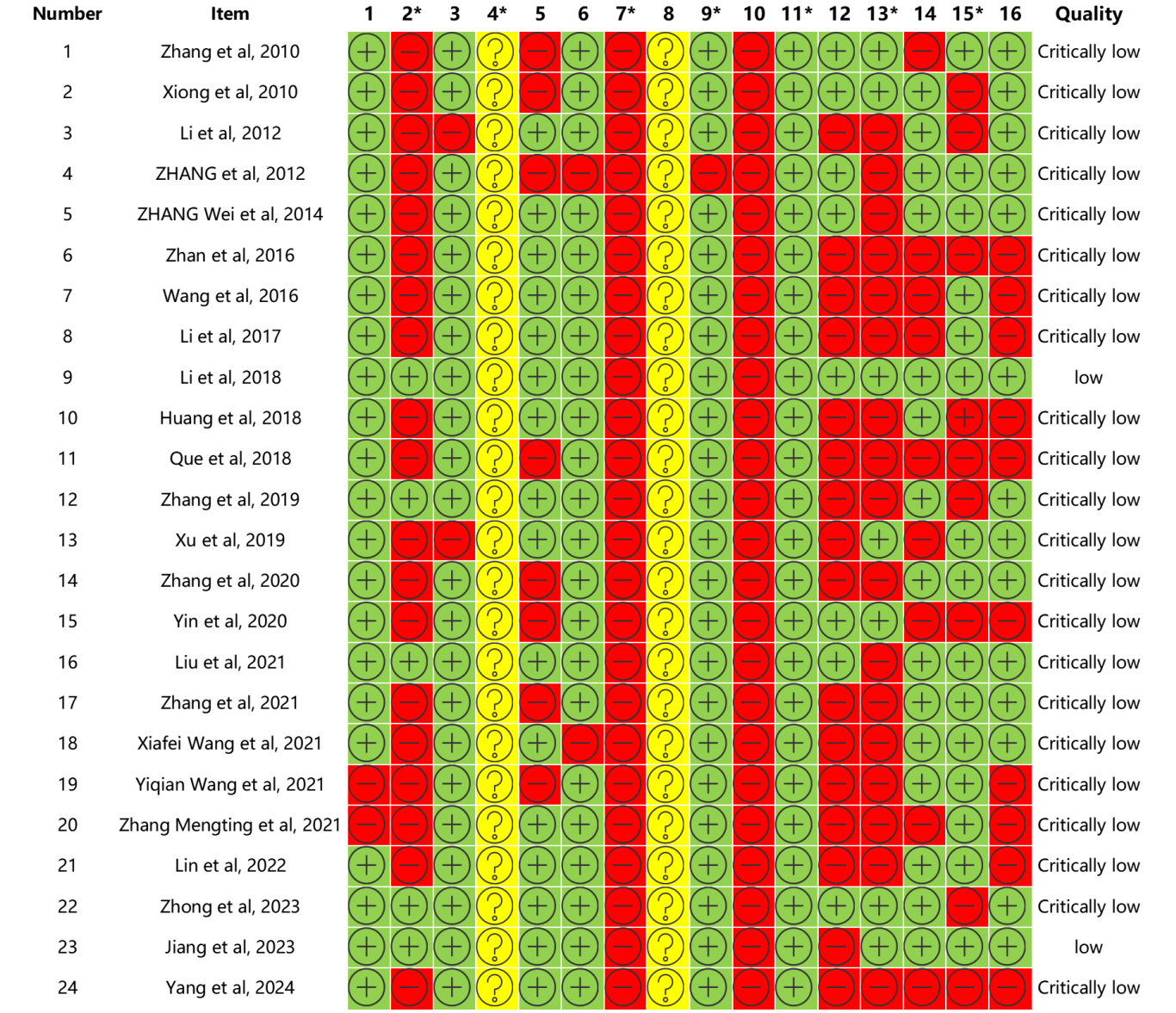


Reporting Quality Appraisal

Reporting quality was evaluated using the PRISMA 2020 checklist. Scores for the 24 studies ranged from 14 to 23. One study [22] was classified as "severely deficient" (score <15), 21 studies [19-21,23-33,35-41] as "somewhat deficient" (score 15 - 21), and 2 studies [34,40] as "relatively complete" (>21). Underreported items included data items (10), methods for evaluating quality of evidence for each outcome (15), presentation of evidence grading (22), registration and protocol (24), competing interests (26), and public information reporting (27) (Figure 3).

XSL•FO
RenderX

Figure 3. Results of the reporting quality evaluation of Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA). "+" denotes yes, "–" denotes no, and "?" denotes partial yes.

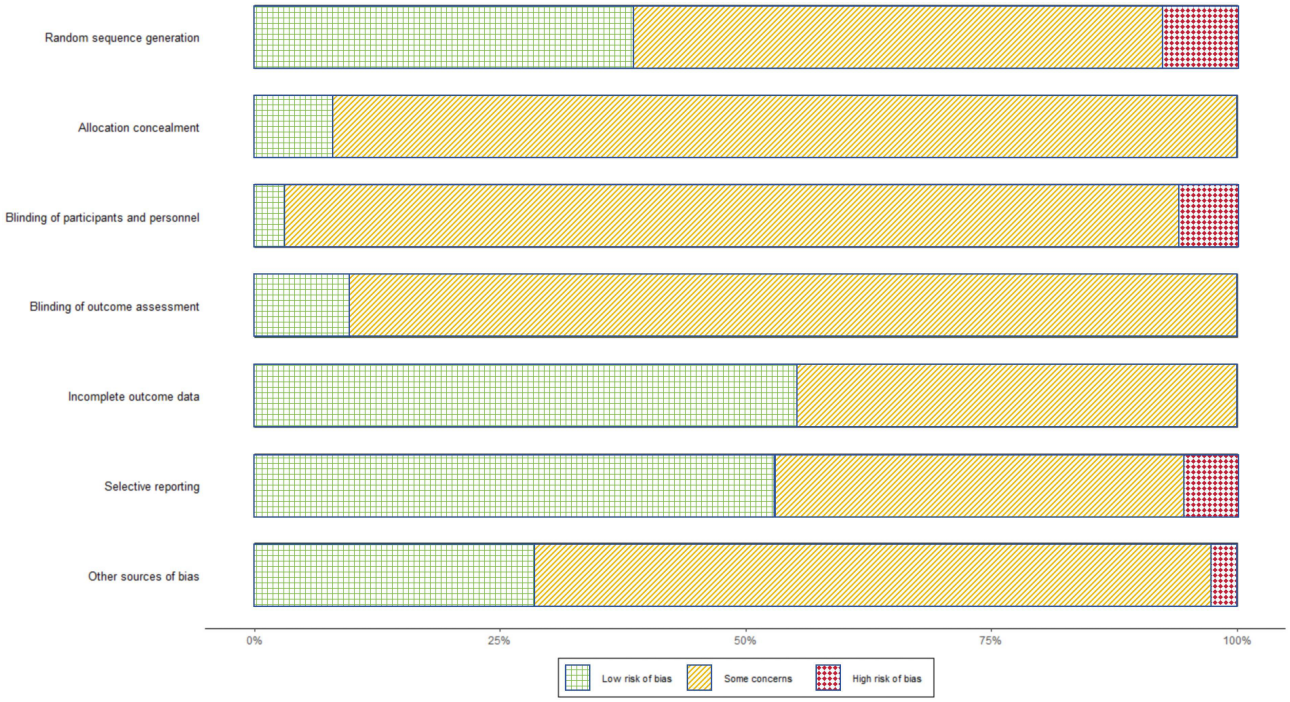


Risk of Bias in Original Literature

After deduplication, 222 original studies remained. Of these, 46 articles (20.7%) were classified as high risk of bias, while the remainder were labeled as having "some concern." Among the high-risk domains, random sequence generation constituted 8.6% (19/222), blinding of participants and personnel 7.7% (17/222), selective reporting 4.5% (10/222), and other biases 1.4% (3/222) (Figure 4). Further details can be found in Multimedia Appendix 2. in

XSL•FO
RenderX

Figure 4. Summary chart of bias risks in original literature. Green squares represent low risk, yellow stripes represent some concerns, and red diamonds represent high risk.



Classification of Evidence Quality

The quality of evidence for the outcome indicators, covering 6 different intervention types, ranged from very low to high. Common downgrading factors included high risk of bias, significant heterogeneity, small sample sizes, and inconsistency of results, as outlined in Table 2 [16,19-41].

XSL•FO
RenderX

Table . Quality of evidence in included systematic reviews with Grading of Recommendations Assessment, Development, and Evaluation.

Studies and outcomes	Number of outcome studies	Publication bias	Risk of bias	Imprecision	Inconsistency	Indirectness	Quality of evidence
Zhang et al (2010) [19]							
Clinical response	13	−1[a]	−1[b]	0	0	0	Low
Reducing depression severity	14	0	−1[b]	0	−1[c]	0	Low
Xiong et al (2010) [20]							
24 HAMD[d] score reduction rates	5	0	0	0	0	0	High
24-HAMD	9	0	−1[b]	0	−1[c]	0	Low
SDS[e]	5	0	0	0	−2[f]	0	Low
Li et al (2012) [21]							
HAMD	8	0	−1[b]	0	0	0	Moderate
Zhang et al (2012) [22]							
Curative rate	15	0	−1[b]	0	0	0	Moderate
Effective rate	15	0	−1[b]	0	0	0	Moderate
Zhang et al (2014) [23]							
HAMD	12	0	−1[b]	0	0	−1[g]	Low
Adverse events	8	0	−1[b]	−1[h]	0	0	Low
Tan et al (2016) [24]							
HAMD	12	−1[a]	−1[b]	0	−2[f]	0	Very low
Effective rate	10	−1[a]	−1[b]	0	0	0	Low
Wang et al (2016) [25]							
HAMD	18	−1[a]	0	0	−1[c]	0	Low
BI[i]	4	0	0	−1[h]	−2[f]	0	Very Low
Li et al (2017) [26]							
HAMD	11	0	−1[b]	0	−2[f]	0	Very Low
Curative rate	8	0	−1[b]	0	0	0	Moderate
Effective rate	9	0	−1[b]	0	0	0	Moderate
Adverse events	2	0	0	−1[h]	0	0	Low
Li et al (2018) [27]							
HAMD	15	0	0	0	0	0	High
Huang et al (2018) [28]							
HAMD	13	0	−1[b]	−1[h]	−1[c]	0	Very low
Que et al (2018) [29]							
Effective rate	11	0	−1[b]	0	0	0	Moderate
HAMD-17	5	0	0	−1[h]	−2[f]	0	Very low
HAMD-24	7	0	−1[b]	0	−2[f]	0	Very low

XSL•FO
RenderX

Studies and outcomes	Number of outcome studies	Publication bias	Risk of bias	Imprecision	Inconsistency	Indirectness	Quality of evidence
Adverse events	9	0	−1[b]	0	0	0	Moderate
Zhang et al (2019) [30]							
Effective rate	7	0	−1[b]	0	0	0	Moderate
Chen et al (2019) [31]							
HAMD	19	0	0	0	−2[f]	0	Low
BI	4	0	0	−1[h]	−2[f]	0	Very low
Effective rate	12	0	0	0	0	0	High
Adverse events	5	0	0	−1[h]	0	0	Moderate
Zhang et al (2020) [32]							
HAMD	13	−1[a]	−1[b]	0	0	0	Low
Effective rate	6	0	−1[b]	−1[h]	0	0	Low
NIHSS[j]	4	0	−1[b]	−1[h]	−1[c]	0	Very low
BI	3	0	−1[b]	−1[h]	0	0	Low
Xiao et al (2020) [33]							
Effective rate	15	0	−1[b]	0	0	0	Moderate
HAMD	15	0	−1[b]	0	−2[f]	0	Very low
Liu et al (2021) [34]							
HAM-D17	8	0	−1[b]	−1[h]	0	0	Low
HAM-D24	4	0	−1[b]	−1[h]	−1[c]	0	Very low
HAMD	6	0	−1[b]	−1[h]	0	0	Low
Adverse events	7	0	−1[b]	v1[h]	0	-1[g]	Very low
Zhang et al (2021) [35]							
HAMD	12	0	−1[b]	0	−2[f]	0	Very low
NIHSS	2	0	−1[b]	−1[h]	0	0	Low
TESS[k]	3	0	−1[b]	−1[h]	−1[c]	0	Very low
Effective rate	12	−1[a]	−1[b]	0	0	0	Low
Wang et al (2021) [36]							
HAMD	19	0	−1[b]	−1[h]	0	0	Low
Adverse events	9	0	−1[b]	0	0	0	Moderate
Wang et al (2021) [37]							
Effective rate	7	−1[a]	0	0	0	0	Moderate
HAMD	7	0	0	0	−1[c]	0	Moderate
Zhang et al (2021) [38]							
HAM-D24	10	0	−1[b]	0	0	0	Moderate
HAM-D17	3	0	0	−1[h]	−1[c]	0	Low

XSL•FO
RenderX

Studies and outcomes	Number of outcome studies	Publication bias	Risk of bias	Imprecision	Inconsistency	Indirectness	Quality of evidence
HAMD score reduction rates	11	0	−1[b]	0	0	0	Moderate
NIHSS	6	0	−1[b]	0	−2[f]	0	Very low
BI	9	0	−1[b]	0	−1[c]	0	Low
Lin et al (2022) [39]							
HAMD	15	0	−1[b]	0	−2[f]	0	Very low
SDS	2	0	0	−1[h]	−1[c]	0	Low
NIHSS	2	0	0	−1[h]	−2[f]	0	Very low
Effective rate	9	0	−1[b]	0	0	0	Moderate
Zhong et al (2023) [16]							
Effective rate	8	0	−1[b]	0	0	0	Moderate
HAMD	7	0	−1[b]	0	−2[f]	0	Very low
SDS	3	0	0	0	−2[f]	0	Low
Jiang et al (2023) [40]							
Effective rate	12	−1[a]	−1[b]	0	0	0	Low
HAMD	10	0	−1[b]	0	−2[f]	0	Very low
Neurological deficit score	6	0	−1[b]	−1[h]	−2[f]	0	Very low
Adverse events	5	0	−1[b]	0	0	0	Moderate
Yang et al (2024) [41]							
HAMD	13	0	−1[b]	0	−2[f]	0	Very low
NIHSS	2	0	0	−1[h]	−2[f]	0	Very low
SDS	B3	0	−1[b]	−1[h]	−2[f]	0	Very low

[a]Asymmetric funnel plots or all positive results may have a large publication bias.

[b]Risk of bias in included studies with respect to randomization, blinding, allocation concealment, completeness of outcome data, or selective reporting risk of bias.

[c]Included in the study 50%≤I^2<75%.

[d]HAMD: Hamilton Depression Rating Scale.

[e]SDS: Self-Rating Depression Scale.

[f]Included in the study 75%≤I^2<100%.

[g]Differences in stroke courses and depression degree exist in the included research.

[h]Inclusion of too small a sample size for the study (sample size for continuous variables <400).

[i]BI: Barthel index.

[j]NIHSS: National Institutes of Health Stroke Scale.

[k]TESS: Treatment Emergent Symptom Scale.

Heterogeneity Assessment

Of the reported outcomes, 25 of 60 (42%) demonstrated no heterogeneity (I^2<25%), 11 of 60 (18%) showed low heterogeneity (25% ≤I^2<50%), 8 of 60 (13%) moderate heterogeneity (50% ≤I^2<75%), and 16 of 60 (27%) high heterogeneity (I^2≥75%). Meta-analyses with high heterogeneity primarily used clinical effective rate and HAMD score as outcomes (Table S2 in Multimedia Appendix 1) [16,19-41].

Overlap of Primary Studies

The primary research indicators we observed were mostly mild to moderate overlap (Table 3). A CCA <10% may suggest incomplete coverage of original studies in the field, with key clinical trials not being adequately included (eg, publication bias of negative results), and a lack of targeted evidence for subgroup populations (eg, different stroke types and severity of depression).

XSL•FO
RenderX

Table . Level of overlap for primary studies among the included systematic reviews.

Outcomes	N[a]	r[b]	c[c]	CCA[d] (%)	Interpretation of overlap
HAMD-24[e]	30	25	4	6.7	Moderate overlap
HAMD-17	15	13	3	7.7	Moderate overlap
HAMD	223	153	18	3	Mild overlap
24 HAMD score reduction rates	17	17	2	0	Mild overlap
Clinical response	13	13	1	0	Mild overlap
Effectiveness rate	121	93	12	2.7	Mild overlap
Curative rate	23	23	2	0	Mild overlap
Barthel index	21	19	3	5.3	Moderate overlap
Treatment Emergent Symptom Scale	3	3	1	0	Mild overlap
SDS[f]	13	12	4	2.8	Mild overlap
NIHSS[g]	17	16	5	1.6	Mild overlap
Barthel index	3	3	1	0	Mild overlap
Adverse events	53	39	8	5.1	Moderate overlap

[a]N: the total number of included observations from all reviews.

[b]r: the number of unique primary studies.

[c]c: the number of systematic reviews.

[d]CCA: corrected covered area.

[e]HAMD: Hamilton Depression Rating Scale.

[f]SDS: Self-Rating Depression Scale.

[g]NIHSS: National Institutes of Health Stroke Scale.

Discussion

Main Findings

Acupuncture therapy is increasingly used for PSD in clinical practice, primarily due to its capacity to mitigate depressive symptoms, enhance patients' quality of life, and diminish the adverse side effects typically associated with pharmacological treatments. In recent years, the body of systematic reviews on acupuncture for PSD has grown substantially. This study undertook a comprehensive evaluation of the existing literature, focusing on methodological quality, reporting quality, and evidence quality, to elucidate current weaknesses and provide guidance for future research. Overall, our findings indicate that the included systematic reviews and meta-analyses generally experienced low methodological, reporting, and evidence quality, largely a result of the poor quality of primary studies, especially due to high risks of bias such as inadequate randomization and insufficient allocation concealment. Furthermore, the primary research indices demonstrated mainly mild to moderate overlap, suggesting that coverage of original studies in this field remains incomplete, leaving potentially pivotal clinical trials unincorporated.

Potential Mechanism

The pathogenesis of PSD is multifactorial. Two major hypotheses involve neurotransmitter imbalance and glutamate-mediated excitotoxicity [50]. The monoaminergic mechanism posits that lower levels of monoamines—such as serotonin, norepinephrine, and dopamine—are associated with depressive symptoms [51]. In addition, glutamate, the brain's principal excitatory neurotransmitter, is often elevated in patients with PSD, reflecting a state of excitotoxicity and contributing to depressive pathology [52].

Acupuncture is believed to modulate these neurochemical disturbances. Animal and clinical studies suggest that acupuncture elevates levels of serotonin, norepinephrine, and DA in the brain, normalizes glutamate concentrations, and upregulates key neurotransmitter receptors with fewer desensitization issues compared with pharmacotherapy [53-55]. Beyond neurotransmitter effects, acupuncture exerts anti-inflammatory actions, reducing levels of proinflammatory cytokines and improving immunological balance—processes increasingly recognized as pivotal in PSD progression [56-61]. Furthermore, acupuncture appears to modulate the hypothalamic–pituitary–adrenal axis, helping to normalize neuroendocrine function, and may also reduce neuronal apoptosis, particularly within the hippocampus, which is closely linked to mood regulation [62-65].

Quality Summaries

The methodological quality of the included reviews was evaluated using the AMSTAR 2 system. The methodological strengths of the included reviews were the use of appropriate methods for outcome statistics (24/24, 100%), a satisfactory

XSL•FO
RenderX

assessment of risk of bias (24/24, 100%), and study selection and data extraction in duplicate (15/24, 62.5%). Common methodological weaknesses included the failure to provide preregistered protocols (5/24, 20.8%), explain ROB in individual studies when interpreting results (7/24, 29.2%), provide a full list of excluded studies (0/24, 0%), and extract data on funding sources (0/24, 0%).

The methodological quality of reviews was systematically assessed using AMSTAR 2. All included reviews applied appropriate statistical methods and risk of bias assessments (24/24, 100%) and most performed duplicate selection and data extraction (15/24, 62.5%). However, major shortcomings were evident: only a minority preregistered protocols (5/24, 20.8%), few explained the risk of bias in interpreting results (7/24, 29.2%), none provided a list of excluded studies (0/24), and none extracted funding source data (0/24). The absence of protocol registration and transparency in study selection raises concerns about selective reporting.

According to the GRADE system, although acupuncture appeared more effective than controls for PSD, none of the 60 outcome variables achieved a high-quality evidence rating, and only 8 were rated as moderate quality. Most underlying RCTs were small, single-center studies, with risk of bias, mostly from inadequate blinding, being the most prominent factor for downgrading.

Regarding PRISMA 2020 criteria, the reporting quality was generally suboptimal: only 1 review [22] had major defects, 21 had certain deficiencies, and 2 [16,34] met standards for completeness. Notably, most reviews neglected protocol registration, subgroup or sensitivity analyses for heterogeneity, or evidence grading for outcomes. Several lacked key visualizations such as flow or bias charts, and some failed to report funding or conflicts of interest, compromising the transparency and objectivity of their results.

Outlook and Recommendations

Almost all of the included SRs exhibited low or very low methodological quality, indicating an urgent need to improve the conduct and reporting of these reviews. Acupuncture shows promise as a complementary approach for improving depressive symptoms after stroke. However, these findings highlight a pressing need for large, well-designed, randomized controlled trials and higher-quality systematic reviews. Assessments typically rely on scales such as HAMD, SDS, and the modified Edinburgh-Scandinavian stroke scale, but these are somewhat subjective and may introduce bias. Future research should emphasize blinding (of both patients and outcome assessors), clearer methodology, robust protocol registration, and transparency regarding funding and conflicts of interest, in order to improve the reliability and applicability of the evidence. In addition, the results showed that there was mainly mild to moderate overlap, and key clinical trials may not have been included. Follow-up studies need to increase the coverage of clinical trials and include multiple languages and countries.

Strengths and Limitations

A principal strength of this review lies in its focus on meta-analyses of RCTs, representing the highest tier of evidence, with rigorous eligibility criteria to minimize confounding. Nonetheless, several limitations should be acknowledged: (1) interventions were heterogeneous across studies, precluding quantitative synthesis; (2) the included systematic reviews generally exhibited low methodological, reporting, and evidence quality; (3) the lack of differentiation between different subtypes of PSD and TCM syndromes complicates the interpretation of the evidence for acupuncture in treating PSD; and (4) the subjective nature of AMSTAR 2 and PRISMA assessments could introduce rater bias.

Conclusions

Overall, the preponderance of current systematic reviews suggests a positive effect of acupuncture in the management of PSD. However, the evidence base is undermined by pervasive deficiencies in methodological rigor and quality of reporting. To advance the field and inform clinical decision-making, further high-quality multicenter RCTs and rigorously conducted systematic reviews are imperative.

Acknowledgments

The authors would like to thank the authors of the primary studies, which were used as a source of information to conduct this study. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

This study received funding from the Regional Key R&D Program of Ningxia Hui Autonomous Region (grant number 2024BBF01004).

Data Availability

All relevant data generated and analyzed are included in this study.

Authors' Contributions

XCT conceived the study. LM and CLX performed the literature search. LM and CLX performed the parsing of the literature. LM and XXH evaluated the methodological quality, reporting quality, and evidence quality of the literature. LM and CLX performed the writing and prepared the figures. XCT edited and revised the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategies for each database and main outcomes and evidence quality.
[DOCX File, 94 KB - i-jmr_v14i1e76577_app1.docx]

Multimedia Appendix 2
Full details of references.
[XLSX File, 25 KB - i-jmr_v14i1e76577_app2.xlsx]

Checklist 1
PRISMA checklist.
[PDF File, 99 KB - i-jmr_v14i1e76577_app3.pdf]

References

1. Medeiros GC, Roy D, Kontos N, Beach SR. Post-stroke depression: a 2020 updated review. Gen Hosp Psychiatry 2020;66:70-80. [doi: 10.1016/j.genhosppsych.2020.06.011] [Medline: 32717644]
2. Chun HYY, Ford A, Kutlubaev MA, Almeida OP, Mead GE. Depression, anxiety, and suicide after stroke: a narrative review of the best available evidence. Stroke 2022 Apr;53(4):1402-1410. [doi: 10.1161/STROKEAHA.121.035499] [Medline: 34865508]
3. Ordatii N, Ordatii A. Correlation of pain and psychological disorder. PMJUA 2020;5(3):12-17. [doi: 10.31636/pmjua.v5i3.2]
4. Zhan Q, Kong F. Mechanisms associated with post-stroke depression and pharmacologic therapy. Front Neurol 2023;14:1274709. [doi: 10.3389/fneur.2023.1274709] [Medline: 38020612]
5. Zhou H, Wei YJ, Xie GY. Research progress on post-stroke depression. Exp Neurol 2024 Mar;373:114660. [doi: 10.1016/j.expneurol.2023.114660] [Medline: 38141804]
6. Masuccio FG, Grange E, Di Giovanni R, Rolla M, Solaro CM. Post-stroke depression in older adults: an overview. Drugs Aging 2024 Apr;41(4):303-318. [doi: 10.1007/s40266-024-01104-1] [Medline: 38396311]
7. Raggi A, Serretti A, Ferri R. A comprehensive overview of post-stroke depression treatment options. Int Clin Psychopharmacol 2024 May 1;39(3):127-138. [doi: 10.1097/YIC.0000000000000532] [Medline: 38170802]
8. Anderson HD, Pace WD, Libby AM, West DR, Valuck RJ. Rates of 5 common antidepressant side effects among new adult and adolescent cases of depression: a retrospective US claims study. Clin Ther 2012 Jan;34(1):113-123. [doi: 10.1016/j.clinthera.2011.11.024] [Medline: 22177545]
9. Dodd S, Mitchell PB, Bauer M, et al. Monitoring for antidepressant-associated adverse events in the treatment of patients with major depressive disorder: an international consensus statement. World J Biol Psychiatry 2018 Aug;19(5):330-348. [doi: 10.1080/15622975.2017.1379609] [Medline: 28984491]
10. Russo NW, Petrucci G, Rocca B. Aspirin, stroke and drug-drug interactions. Vascul Pharmacol 2016 Dec;87:14-22. [doi: 10.1016/j.vph.2016.10.006] [Medline: 27765537]
11. Cai W, Ma W, Li YJ, Wang GT, Yang H, Shen WD. Efficacy and safety of electroacupuncture for post-stroke depression: a randomized controlled trial. Acupunct Med 2022 Oct;40(5):434-442. [doi: 10.1177/09645284221077104] [Medline: 35232229]
12. Hu X, Li J. Acupuncture combined with repeated transcranial magnetic stimulation for mild to moderate post-stroke depression: a randomized controlled trial [Article in Chinese]. Zhongguo Zhen Jiu 2024 Jul 12;44(7):765-769. [doi: 10.13703/j.0255-2930.20230908-0007] [Medline: 38986588]
13. Yang NN, Lin LL, Li YJ, et al. Potential mechanisms and clinical effectiveness of acupuncture in depression. Curr Neuropharmacol 2022;20(4):738-750. [doi: 10.2174/1570159X19666210609162809] [Medline: 35168522]
14. You Y, Zhang T, Shu S, Qian X, Zhou S, Yao F. Wrist-ankle acupuncture and Fluoxetine in the treatment of post-stroke depression: a randomized controlled clinical trial. J Tradit Chin Med 2020 Jun;40(3):455-460. [doi: 10.19852/j.cnki.jtcm.2020.03.014] [Medline: 32506860]
15. Zhang L, Zhong Y, Quan S, et al. Acupuncture combined with auricular point sticking therapy for post stroke depression:a randomized controlled trial [Article in Chinese]. Zhongguo Zhen Jiu 2017 Jun 12;37(6):581-585. [doi: 10.13703/j.0255-2930.2017.06.003] [Medline: 29231496]
16. Zhong D, Cheng H, Pan Z, et al. Efficacy of scalp acupuncture combined with conventional therapy in the intervention of post-stroke depression: a systematic review and meta-analysis. Complement Ther Med 2023 Oct;77:102975. [doi: 10.1016/j.ctim.2023.102975] [Medline: 37619716]
17. Wang X, Xiong J, Yang J, et al. Meta-analysis of the clinical effectiveness of combined acupuncture and Western medicine to treat post-stroke depression. J Tradit Chin Med 2021 Feb;41(1):6-16. [doi: 10.19852/j.cnki.jtcm.2021.01.002] [Medline: 33522192]

XSL•FO
RenderX

18. 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]
19. Zhang ZJ, Chen HY, Yip KC, Ng R, Wong VT. The effectiveness and safety of acupuncture therapy in depressive disorders: systematic review and meta-analysis. J Affect Disord 2010 Jul;124(1-2):9-21. [doi: 10.1016/j.jad.2009.07.005] [Medline: 19632725]
20. Xiong J, Du YH, Liu JL, et al. Acupuncture versus Western medicine for post stroke depression: a systematic review [Article in Chinese]. J Evid Based Med 2010;10(3):179-185. [doi: 10.3969/j.issn.1671-5144.2010.03.021]
21. Lai SW, Chen J, Hu J, Wang H, Wang S. Meta-analysis comparing the effects of combined acupuncture and medicine with Western medicine on improving depressive status in patients with post-stroke depression [Article in Chinese]. Shaanxi Traditional Chinese Medicine 2012;33(9):1263-1267 [FREE Full text]
22. Zhang GC, Fu WB, Xu NG, et al. Meta analysis of the curative effect of acupuncture on post-stroke depression. J Tradit Chin Med 2012 Mar;32(1):6-11. [doi: 10.1016/s0254-6272(12)60024-7] [Medline: 22594095]
23. Zhang W, Sun JH, Gao Y, et al. System review on treating post-stroke depression with acupuncture. World J Acupunct Moxibustion 2014 Jun;24(2):52-59. [doi: 10.1016/S1003-5257(14)60026-X]
24. Tan F, Cheng NF, Tan JQ. Systematic evaluation of the efficacy of electroacupuncture and antidepressant western medicine in treating post-stroke depression [Article in Chinese]. J Tradit Chin Med 2016;34(10):2379-2383 [FREE Full text]
25. Wang Q, Zheng S, Tang M, Cui H. Meta-analysis of body acupuncture treatment for post-stroke depression [Article in Chinese]. Journal of Traditional Chinese Medicine 2016;31(8):1227-1231 [FREE Full text]
26. Li W, Ding SN, Zhang JG. Systematic evaluation of the effects of two methods of treating depressed patients. Journal of Huzhou Normal 2017;39(10):84-91 [FREE Full text]
27. Li XB, Wang J, Xu AD, et al. Clinical effects and safety of electroacupuncture for the treatment of post-stroke depression: a systematic review and meta-analysis of randomised controlled trials. Acupunct Med 2018 Oct;36(5):284-293. [doi: 10.1136/acupmed-2016-011300] [Medline: 29776950]
28. Huang YJ, Li LX, Zhou YL, Wu J, Lin GH, Ming KW. Meta-analysis of the efficacy of acupuncture and electroacupuncture for post-stroke depression. Mod Hosp 2018;18(1):120-124 [FREE Full text]
29. Que F, Wen Y, Wang W. A systematic review of the efficacy of acupuncture compared with Prozac in the treatment of post-stroke depression [Article in Chinese]. Hunan Journal of Traditional Chinese Medicine 2018;34(9):141-145 [FREE Full text]
30. Zhang XY, Li YX, Liu DL, Zhang BY, Chen DM. The effectiveness of acupuncture therapy in patients with post-stroke depression: an updated meta-analysis of randomized controlled trials. Medicine (Baltimore) 2019 May;98(22):e15894. [doi: 10.1097/MD.0000000000015894] [Medline: 31145349]
31. Xu R, Chen J. Meta-analysis and trial sequential analysis of acupuncture treatment for depression in patients with post-stroke depression [Article in Chinese]. Chin J Gerontol 2019;39(23):5722-5726 [FREE Full text]
32. Zhang K, Cui G, Gao Y, Shen W. Does acupuncture combined with antidepressants have a better therapeutic effect on post-stroke depression? A systematic review and meta-analysis. Acupunct Med 2021 Oct;39(5):432-440. [doi: 10.1177/0964528420967675] [Medline: 33334118]
33. Yin M, Wei X, Liang F. Meta-analysis of acupuncture treatment for post-stroke depression [Article in Chinese]. Henan Traditional Chinese Medicine 2020;40(11):1729-1732 [FREE Full text]
34. Liu R, Zhang K, Tong QY, Cui GW, Ma W, Shen WD. Acupuncture for post-stroke depression: a systematic review and meta-analysis. BMC Complement Med Ther 2021 Apr 1;21(1):109. [doi: 10.1186/s12906-021-03277-3] [Medline: 33794857]
35. Zhang L, Chen B, Yao Q, et al. Comparison between acupuncture and antidepressant therapy for the treatment of poststroke depression. Medicine (Abingdon) 2021;100(22):e25950. [doi: 10.1097/MD.0000000000025950] [Medline: 34087835]
36. Wang X, Cai W, Wang Y, Huang S, Zhang Q, Wang F. Is electroacupuncture an effective and safe treatment for poststroke depression? An updated systematic review and meta-analysis. Biomed Res Int 2021;2021(1):8661162. [doi: 10.1155/2021/8661162] [Medline: 34485527]
37. Wang Y, Yuan C, Yang J, Li J, Huang H. A systematic review of the efficacy and safety of acupuncture in treating post-stroke depression [Article in Chinese]. Shandong Journal of Traditional Chinese Medicine 2021;40(5):464-470 [FREE Full text]
38. Zhang MT, Tan K, Hu S, Liao LY, Shi WY, Zhou WJ. Meta-analysis of acupuncture to improve the depressive state and daily life ability of post-stroke depressed patients [Article in Chinese]. Shandong J Tradit Chin Med 2021;55(1):13-19. [Medline: 34392647]
39. Lin S, Lu C, Zhuo Y, Yuan J, Zhu Y, Pi M. Meta-analysis of the efficacy and safety of brain-awakening acupuncture in the treatment of post-stroke depression [Article in Chinese]. Journal of Guangzhou University of Chinese Medicine 2022;39(6):1453-1460 [FREE Full text]
40. Jiang W, Jiang X, Yu T, Gao Y, Sun Y. Efficacy and safety of scalp acupuncture for poststroke depression: a meta-analysis and systematic review. Medicine (Baltimore) 2023 Aug 4;102(31):e34561. [doi: 10.1097/MD.0000000000034561] [Medline: 37543780]
41. Yang Y, Cheng Z, Xu W, Ye W, Xiong J, Hong E. Meta-analysis of the effectiveness and safety of electroacupuncture in treating post-stroke depression [Article in Chinese]. Modern Distance Education of Traditional Chinese Medicine 2024;22(8) [FREE Full text]

XSL•FO
RenderX

42. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008. [doi: 10.1136/bmj.j4008] [Medline: 28935701]
43. Zhou WW, Xu JF, Shi XT, et al. Quality evaluation on the systematic reviews/meta-analyses related to interventions published in the Chinese Journal of Evidence-based Medicine. Chin J Evid Based Med 2013;13(4):482-488. [doi: 10.7507/1672-2531.20130082]
44. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008 Apr 26;336(7650):924-926. [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
45. Belbasis L, Bellou V, Ioannidis JPA. Conducting umbrella reviews. BMJ Med 2022;1(1):e000071. [doi: 10.1136/bmjmed-2021-000071] [Medline: 36936579]
46. Andrews J, Guyatt G, Oxman AD, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol 2013 Jul;66(7):719-725. [doi: 10.1016/j.jclinepi.2012.03.013] [Medline: 23312392]
47. Petitti DB. Approaches to heterogeneity in meta-analysis. Stat Med 2001 Dec 15;20(23):3625-3633. [doi: 10.1002/sim.1091] [Medline: 11746342]
48. Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. J Health Serv Res Policy 2002 Jan;7(1):51-61. [doi: 10.1258/1355819021927674] [Medline: 11822262]
49. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
50. Loubinoux I, Kronenberg G, Endres M, et al. Post‐stroke depression: mechanisms, translation and therapy. J Cellular Molecular Medi 2012 Sep;16(9):1961-1969 [FREE Full text] [doi: 10.1111/j.1582-4934.2012.01555.x] [Medline: 22348642]
51. Dulay MF, Criswell A, Hodics TM. Biological, psychiatric, psychosocial, and cognitive factors of poststroke depression. Int J Environ Res Public Health 2023 Mar 30;20(7):37047944. [doi: 10.3390/ijerph20075328] [Medline: 37047944]
52. Gruenbaum BF, Kutz R, Zlotnik A, Boyko M. Blood glutamate scavenging as a novel glutamate-based therapeutic approach for post-stroke depression. Ther Adv Psychopharmacol 2020;10:2045125320903951. [doi: 10.1177/2045125320903951] [Medline: 32110376]
53. Qin Y, Dong H, Sun Y, Cheng X, Yao H. Effects of different acupuncture schemes on neurotransmitters and related inflammatory factors in rats with post-stroke depression [Article in Chinese]. Chinese Journal of Rehabilitation Theory and Practice 2023;29(1):30-37 [FREE Full text]
54. Cui Y, Wei F, Yan W, Song Y, Shi SJ, Wang HP. Influencing factors of depression onset after stroke [Article in Chinese]. Shandong Medicine 2017;57(9):89-91 [FREE Full text]
55. Xu N, He Y, Wei YN, Bai L, Wang L. Possible antidepressant mechanism of acupuncture: targeting neuroplasticity. Front Neurosci 2025;19:1512073. [doi: 10.3389/fnins.2025.1512073] [Medline: 40018358]
56. Makris K, Haliassos A, Chondrogianni M, Tsivgoulis G. Blood biomarkers in ischemic stroke: potential role and challenges in clinical practice and research. Crit Rev Clin Lab Sci 2018 Aug;55(5):294-328. [doi: 10.1080/10408363.2018.1461190] [Medline: 29668333]
57. Jiang J, Xie H, Cao S, et al. Post-stroke depression: exploring gut microbiota-mediated barrier dysfunction through immune regulation. Front Immunol 2025;16(1547365):1547365. [doi: 10.3389/fimmu.2025.1547365] [Medline: 40098959]
58. Leonard BE. Inflammation and depression: a causal or coincidental link to the pathophysiology? Acta Neuropsychiatr 2018 Feb;30(1):1-16. [doi: 10.1017/neu.2016.69] [Medline: 28112061]
59. Liu L, Shi L, Leng F. Discussion on the effects of acupuncture therapy on immune factors in post-stroke depression model rats based on p38 MAPK signaling pathway. Modern Journal of Integrated Traditional Chinese and Western Medicine 2022;31(7):914-917 [FREE Full text]
60. Liu Y, Zhang G, Li J, Lv Y, Qi R. Clinical efficacy of acupuncture combined with escitalopram oxalate in the treatment of mild-to-moderate post-stroke depression. Neuropsychiatr Dis Treat 2025;21:917-925. [doi: 10.2147/NDT.S507265] [Medline: 40260315]
61. Zhang Y, Tang YW, Peng YT, Yan Z, Zhou J, Yue ZH. Acupuncture, an effective treatment for post-stroke neurologic dysfunction. Brain Res Bull 2024 Sep;215:111035. [doi: 10.1016/j.brainresbull.2024.111035] [Medline: 39069104]
62. Juruena MF. Early-life stress and HPA axis trigger recurrent adulthood depression. Epilepsy Behav 2014 Sep;38(148-59):148-159. [doi: 10.1016/j.yebeh.2013.10.020] [Medline: 24269030]
63. Zuo Y, Lou B, Liu K, Gao Y. Effect of electro-needling acupoints of pericardium meridian on blood ACTH and Cort in treatment of post-stroke depression [Article in Chinese]. Journal of Clinical Acupuncture 2023;39(4):53-57 [FREE Full text]
64. Zheng J, Wang Y, Zhang C, et al. Electroacupuncture negatively regulates the nesfatin-1/ERK/CREB pathway to alleviate HPA axis hyperactivity and anxiety-like behaviors caused by surgical trauma. Chin Med 2024 Aug 17;19(1):108. [doi: 10.1186/s13020-024-00974-2] [Medline: 39153974]
65. Chen L, Wang L. Effects of “Tongdu Tiaoshen” acupuncture on Bax and Bcl-2 in the hippocampus of rats with depression [Article in Chinese]. Journal of Traditional Chinese Medicine and Pharmacy 2011;39(6):85-88 [FREE Full text]

XSL•FO
RenderX

Abbreviations

AMSTAR 2: Assessment of Multiple Systematic Reviews-2
CCA: corrected covered area
GRADE: Grading of Recommendations Assessment, Development and Evaluation
HAMD: Hamilton Depression Rating Scale
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PSD: poststroke depression
RCT: randomized controlled trial
SDS: Self-Rating Depression Scale

Edited by TDA Cardoso; submitted 26.04.25; peer-reviewed by C Ke, J Hao, Y Zhang; revised version received 14.07.25; accepted 15.07.25; published 16.10.25.

Please cite as:
Meng L, Xu CL, He XX, Tan XC
Acupuncture and Moxibustion for Poststroke Depression: Systematic Review
Interact J Med Res 2025;14:e76577
URL: https://www.i-jmr.org/2025/1/e76577
doi:10.2196/76577

© Lu Meng, Chuang-Long Xu, Xiao-Xu He, Xiao-Chan Tan. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 16.10.2025. 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

Adoption of Telemedicine for Dementia Care in Nigeria: Scoping Review

Abiodun Adedeji[1], PhD; Huseyin Dogan[1], PhD; Festus Adedoyin[1], PhD; Michelle Heward[2], PhD

[1]Department of Computing and Informatics, Faculty of Science and Technology, Bournemouth University, Talbot Campus, Fern Barrow, Poole, Dorset, Bournemouth, United Kingdom

[2]Department of Psychology, Faculty of Science and Technology, Bournemouth University, Bournemouth, United Kingdom

Corresponding Author:
Abiodun Adedeji, PhD
Department of Computing and Informatics, Faculty of Science and Technology, Bournemouth University, Talbot Campus, Fern Barrow, Poole, Dorset, Bournemouth, United Kingdom

Abstract

Background: Dementia is a global health challenge, particularly in Nigeria, where limited health care infrastructure, cultural stigmas, and poor awareness hinder its care. Telemedicine can improve patient outcomes, increase health care access, and support caregivers. However, challenges such as poor internet connectivity, digital literacy, and a lack of integrated strategies hinder its adoption, particularly in rural areas.

Objective: This scoping review aims to evaluate the adoption of telemedicine for dementia care in Nigeria by highlighting existing interventions, their effectiveness, implementation challenges, and contextual barriers. It also draws on global evidence to propose culturally relevant, sustainable strategies.

Methods: A scoping review was conducted using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) framework. Peer-reviewed articles were included if they focused on telemedicine or digital health interventions for dementia care in Nigeria or sub-Saharan Africa and published between January 2010 and February 2024. Databases searched included PubMed, Scopus, CINAHL, PsycINFO, Cochrane Library, and Google Scholar. A total of 23 articles met the inclusion criteria.

Results: Among the 23 studies, 10 (43.5%) focused on mobile health apps, 8 (34.8%) on video consultations, and 5 (21.7%) on remote monitoring tools. These interventions improved caregiver support, medication adherence, and access to specialist care. Key barriers included limited digital literacy, poor internet access, and a lack of cohesive national telemedicine policy.

Conclusions: There is an urgent need for an inclusive national telemedicine policy in Nigeria. Interventions such as mobile health, video consultations, and remote monitoring tools show potential to enhance dementia care, reduce caregiver burden, and improve health outcomes.

(Interact J Med Res 2025;14:e75168) doi:10.2196/75168

KEYWORDS

telemedicine; dementia care; Nigeria; mHealth; video consultations; remote monitoring; caregiver support; health care accessibility; national telemedicine strategy; mobile health

Introduction

Overview

Dementia is a progressive neurodegenerative condition marked by declining memory, cognitive abilities, and daily functioning, and it presents a significant global health burden [1]. According to the World Health Organization, more than 55 million people are currently affected worldwide, with nearly 10 million new cases emerging each year. This number is expected to rise to 74.7 million by 2030 and 131.5 million by 2050, underscoring the urgent need for scalable and sustainable models of care. Although the social and health care implications of dementia are increasingly recognized globally, its impact in culturally diverse and resource-limited settings, such as Nigeria, remains insufficiently examined.

Nigeria, Africa's most populous country, is home to more than 250 ethnic groups and more than 500 languages, yet its adult literacy rate is only 62% with significant urban-rural disparities [2]. The education system faces diglossia, as English contrasts with indigenous languages, affecting health care communication and dementia care. Early-onset dementia is rare but tends to progress more rapidly [2]. Clinical presentations of dementia in Nigeria differ from those in developed countries due to cultural interpretations of aging conditions [3]. Behavioral and psychological symptoms are poorly recognized and

XSL•FO
RenderX

underreported due to stigma and lack of awareness in rural areas (Figure 1).

Figure 1. Geographical map highlighting the northeastern Nigerian region for telemedicine dementia care.



Digital literacy worsens these issues: mobile phone adoption has increased in Nigeria, with more than 70% of the population owning at least 1 device [4]. This highlights the need to address awareness gaps for better telemedicine use in dementia care.

The article evaluates Nigeria's telemedicine adoption for dementia care, examining sociocultural, linguistic, and technological contexts, while identifying gaps to improve outcomes [5].

The Role of Telemedicine in Dementia Care

Telemedicine enhances dementia care in Nigeria by providing remote clinical services. According to Louis et al [6], this system includes diagnosis and consultation via mobile phones and computers [7].

Telemedicine allows patients and caregivers to connect with health care professionals, reducing the psychological impacts of distancing [8]. During the pandemic, remote consultations, monitoring, and cognitive assessments eased isolation burdens while strengthening relationships with families and health care professionals, improving well-being and identity [9,10].

Telemedicine use in Nigeria increased during COVID-19, highlighting the need for remote care as face-to-face became limited [11]. Restrictions worsened social isolation and reduced health care access for individuals living with dementia and caregivers [12]. Telemedicine supports diagnosis, caregiver assistance, and community education [10]. In Nigeria, telemedicine provides a viable means to improve dementia care through remote services using computers and mobile phones [13,14].

Dementia remains a global health concern [15]. Increased awareness has emphasized caregiver support, with telemedicine emerging as a key intervention for COVID-19 [15,16].

This review discusses the effectiveness and challenges of telemedicine interventions for dementia care in Nigeria. It also aims to guide future development in resource-poor settings by assessing current usage [17]. Active interventions such as videoconferencing, remote assessment, and mobile apps are prioritized over passive monitoring technologies. According to Haimi [18], research highlights issues such as poor internet connectivity and technological inequality. Despite these issues, telemedicine shows promising potential for improving dementia care in Nigeria and justifies further studies [19].

Conceptual Framework

The conceptual framework shows interactions between caregivers, health care professionals, individuals living with dementia, and technologies to enhance care delivery [13]. Mobile health (mHealth) app video consultations and monitoring devices improve access and integrate stakeholders into a unified system [20]. Caregivers use telemedicine tools for care management, while health care providers use them for diagnosis, treatment, and follow-up. These tools help overcome geographical barriers [21]. Technologies support caregivers in planning routines and seeking advice [22]. Video consultations address and improve specialist access, and monitoring systems

XSL•FO
RenderX

track health for timely intervention [23]. This framework improves health care delivery, reduces caregiver burden, and provides an improved quality of life [24], helping stakeholders identify areas needing improvement.

Digital Literacy

Digital literacy is essential for telemedicine adoption, especially in resource-poor countries such as Nigeria [25]. It remains a challenge in urban and rural areas [26]. Mobile phone penetration is 70%, but rural digital literacy is about 30% compared to 62% in urban areas, highlighting a digital gap affecting telemedicine [23]. Urban caregivers benefit from better infrastructure, while rural caregivers face limited access, education, and training [27]. Low literacy in rural areas hinders the use of telemedicine tools, such as mHealth or video consultations for specialist care [28,29].

Challenges of Telemedicine Adoption in Nigeria

Despite the potential benefits, barriers hinder telemedicine use in Nigeria, especially in rural areas [30]. Technological challenges include infrastructure, internet services, and digital literacy, while nontechnological ones involve ethical concerns and funding issues [31]. This scoping review evaluates the adoption of telemedicine for dementia care in Nigeria, highlighting interventions, effectiveness, challenges, and contextual barriers. It draws on global evidence to propose culturally relevant and sustainable strategies for implementation.

Methods

Study Design

This study used a scoping review methodology, guided by the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) framework. The purpose of this work was to explore the adoption of telemedicine for dementia care in Nigeria by identifying current interventions, implementation challenges, and research gaps.

Eligibility Criteria

Peer-reviewed articles were eligible for inclusion if they (1) focused on dementia care in Nigeria or sub-Saharan Africa; (2) included individuals living with dementia, caregivers, or health care professionals; (3) reported on telemedicine or digital health interventions (eg, mHealth apps, video consultations, or remote monitoring); (4) were published in English between January 2010 and March 2024; and (5) applied empirical methods, including qualitative, quantitative, or mixed-methods design. Conversely, studies were excluded if they (1) focused on diseases other than dementia, (2) lacked telemedicine components, (3) were purely theoretical or not empirical, or (4) were not published in English (Table 1).

Table . Inclusion criteria and exclusion criteria for the scoping review.

Criteria	Inclusion	Exclusion
Population	Individual living with dementia, caregiver, or health care professionals	Other diseases than dementia
Interventions	Studies that involve telemedicine interventions or telehealth	Studies that do not involve telemedicine interventions
Study design	Studies that include qualitative, quantitative, and mixed designs	Studies that are purely theoretical
Outcomes measure	The effectiveness of telemedicine on dementia care	No outcomes related to the impact of telemedicine on dementia care
Publication date	Studies published from 2010 to 2024	Studies published before 2010
Language	Studies published in English	Studies published in languages other than English

Information Sources

The following databases were searched: PubMed, Scopus, CINAHL, PsycINFO, Cochrane Library, and Google Scholar. In addition, citation searching involved backward citation tracking from the reference lists of included articles and relevant systematic reviews to identify further relevant studies.

Search Strategy

A comprehensive search strategy was developed using keywords such as “telemedicine,” “dementia care,” “digital health,” “remote care,” “caregiver support,” “Nigeria,” and “Sub-Saharan Africa.” Boolean operators and truncations were used to optimize search results. The full search strategy is available in Multimedia Appendix 1.

Study Selection

Study selection followed a 2-stage screening process in accordance with the PRISMA-ScR framework. In the first stage, 640 records were identified through searches of PubMed, Scopus, CINAHL, PsycINFO, Cochrane Library, and Google Scholar. After removing 220 duplicate records, 420 titles and abstracts were screened independently by 2 reviewers (AA and EO). In the second stage, 100 full-text articles were also independently assessed for eligibility by the same 2 reviewers. At both stages (title or abstract and full-text screening), disagreements were resolved through discussion with a third reviewer (HD).

Data Charting Process

Data were independently extracted by 2 reviewers (AA and EO) using a standardized data charting form, which was piloted and

XSL•FO
RenderX

refined prior to full data extraction. Discrepancies were resolved through discussion, and when necessary, by involving a third reviewer (HD).

Data Items

The following items were extracted from each included study: authorship, publication year, country, study design, target population, type of telemedicine intervention, measured outcomes, and main findings.

Synthesis of Results

A narrative synthesis was used to categorize the data into 3 main themes: (1) types of telemedicine interventions, (2) barriers to adoption, and (3) facilitating factors. Where applicable, frequencies and percentages were used to summarize the distribution of study characteristics and intervention types.

Data Synthesis by Intervention Type

The extracted data were grouped by type of telemedicine intervention. These included mHealth apps, video consultations, and remote monitoring tools. A narrative synthesis was conducted to identify patterns across studies and categorize findings into key intervention types. The goal was to compare their reported outcomes, implementation contexts, and barriers in dementia care delivery.

Results

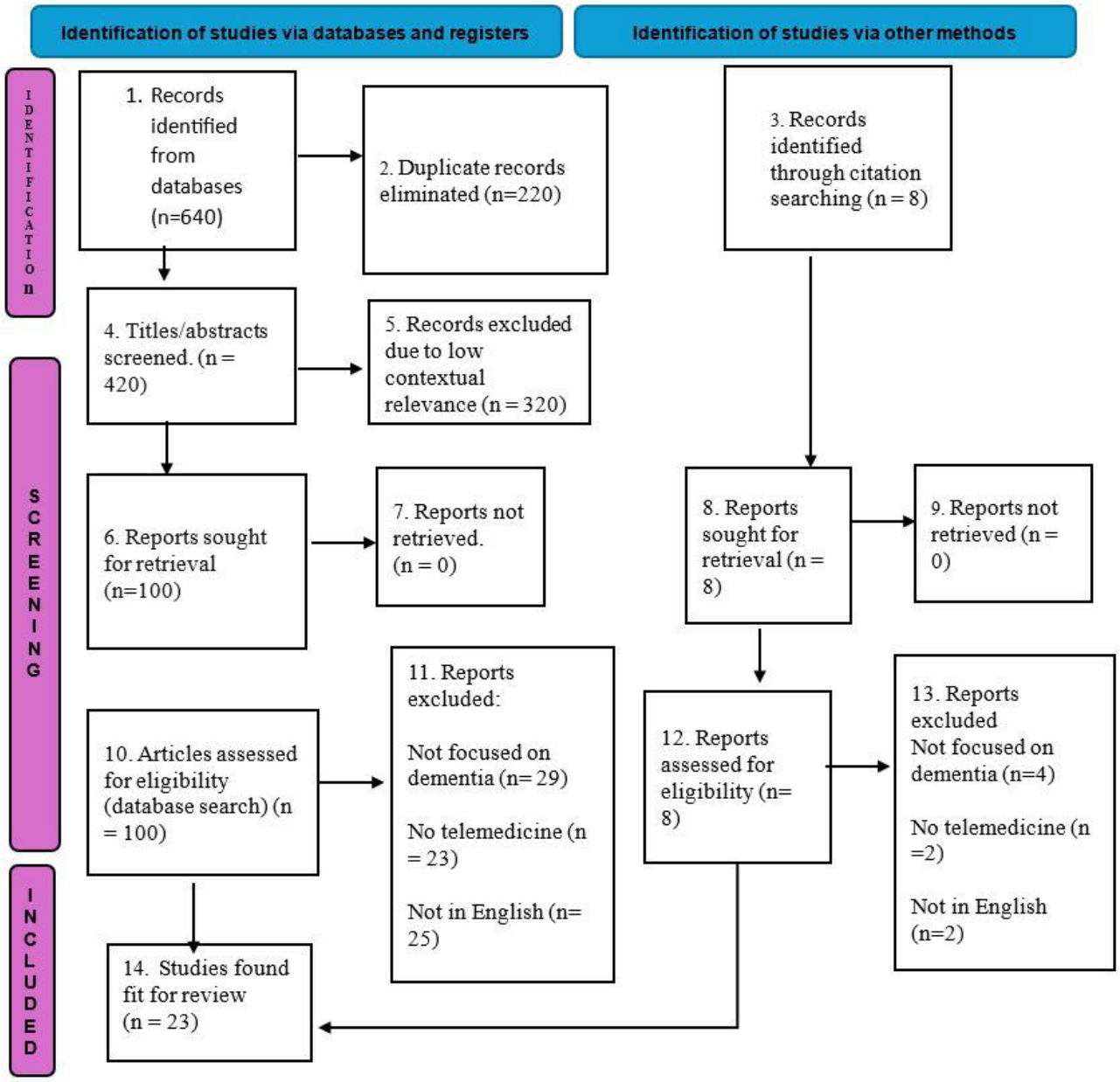
Overview

The reviewed studies explored a variety of telemedicine interventions applied to dementia care within Nigerian or sub-Saharan African contexts. These interventions included mHealth apps, video consultations, and remote patient monitoring systems aimed at supporting caregivers, improving access to specialist care, and enhancing patient outcomes. The included studies used diverse methodological approaches, including qualitative interviews, cross-sectional surveys, randomized controlled trials (RCTs), and mixed methods designs. Terms commonly used across the studies included “telemedicine,” “digital health,” “remote care,” “mHealth,” and “eHealth,” reflecting a broad conceptual framing of digital health delivery in dementia care.

The study selection process is outlined in Figure 2 below. Out of 640 initial records, 420 were screened, 40 full-text articles were assessed, and a total of 23 studies met the inclusion criteria and were included in the final review. An additional 8 records were identified through backward citation searching of included studies and relevant systematic reviews. All 8 records were retrieved and assessed for eligibility. However, none met the inclusion criteria due to reasons such as not focusing on dementia (n=4), lack of telemedicine components (n=2), and not being published in English (n=2).

XSL•FO
RenderX

Figure 2. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram showing the selection process.



Characteristics of Included Studies

In various regions, the primary characteristics of the studies are outlined in Table 2. A total of 23 studies were conducted mainly in Nigeria (n=7) [1,6,9,11,14,32,33], followed by the United States (n=5) [34-38], the United Kingdom (n=6) [39-44], Hong Kong (n=2) [45,46], Greece (n=1) [13], Germany (n=1) [47], and Spain (n=1) [48]. Study settings varied, individuals' homes (n=6), health care institutions (n=10), and mixed environments (n=7).

XSL•FO
RenderX

Table . Characteristics of included studies on telemedicine and dementia care in Nigeria and other sub-Saharan countries.

S.No	Author (year)	Country published	Study design	Intervention type	Sample size (n)	Technology type	Key measure	Outcomes	Modality
1	Muili et al [1] (2023)	Nigeria	Review	Remote consultation, video teleconferencing	—[a]	Cognitive tests, MMSE[b], HVLT-R[c], and letter fluency	Feasibility of telemedicine for dementia diagnosis and follow-up	Internet-based	Synchronous
2	Oyinlola et al [11] (2024)	Nigeria	Mixed method	Remote monitoring	—	Barriers and facilitators of telemedicine	Need for national strategy	Internet-based	Asynchronous
3	Louis et al [6] (2021)	Nigeria	Quantitative	Remote monitoring, mHealth[d]	—	Caregiver satisfaction and system usability	Improved usability lowers stress	Smartphone-based	Asynchronous
4	Ibrahim et al [9] (2024)	Nigeria	Mixed method	Video consultations	—	Pandemic-related care disruption	Persistent access during COVID-19 reduced isolation	Internet-based	Synchronous
5	Angelopoulou et al [13] (2022)	Greece	Narrative review	Remote monitoring	—	Feasibility and care efficiency	Improved quality of care and reduced travel burden	Internet-based	Asynchronous & synchronous
6	Anthony [14] (2021)	Nigeria	Case report	Telehealth for rural dementia care	—	Infrastructure barriers	Need for policy framework integration	Internet-based	Synchronous & asynchronous
7	Strini et al [45] (2023)	Hong Kong	Mixed method	mHealth, video telehealth	70	Caregiver burden QoL[e] metrics	Reduced stress and higher QoL for caregivers	Smartphone-based	Synchronous
8	Page et al [46] (2021)	Hong Kong	Quantitative	Video conferencing for dementia care	60	Cognitive functioning QoL-AD[f] and caregiver burden	Improved resilience and better scores reduced burden	Internet-based	Synchronous
9	Ojeahere et al [34] (2020)	US	Quantitative	Wearable sensors for dementia monitoring	50	Daily activity tracking for early detection	Improved monitoring of timely interventions	Wearable sensor	Asynchronous
10	Cote et al [39] (2020)	UK	Qualitative	Remote monitoring tools	—	Daily activity tracking and early detection	Improved monitoring of timely interventions	Wearable sensor	Asynchronous
11	Gaugler et al [35] (2021)	US	Systematic review	Care planning, video consultations	250	Cognitive status and care planning	Improved dementia care planning and QoL	Internet-based	Synchronous

XSL•FO
RenderX

S.No	Author (year)	Country published	Study design	Intervention type	Sample size (n)	Technology type	Key measure	Outcomes	Modality
12	Yi et al [36] (2021)	US	Qualitative	mHealth for symptom tracking	40	Behavioral symptoms and medication adherence	Improved monitoring reduced hospital visits	Smartphone-based	Asynchronous & Synchronous
13	Mason et al [37] (2022)	US	Cross-sectional	Video telehealth for veterans	24	Semi-structured interviews	Reduced travel burden for caregivers	Internet-based	Synchronous
14	Gately et al [32] (2019)	Nigeria	Qualitative	Telemedicine in hospitals	7	Implementation issues	Improved medication adherence and caregiver satisfaction	Internet-based	Synchronous
15	Adenuga et al [33] (2020)	Nigeria	Review	mHealth interventions	—	App features and AI integration	Improved caregiver support and usability	Smartphone-based	Asynchronous
16	Deniz-Garcia et al [40] (2023)	UK	Scoping review	Implementation barriers	—	Cost access and policy limitations	Policy recommendations	Internet-based	Asynchronous
17	Arora et al [38] (2024)	US	Systematic review	Telemedicine platforms	—	Cognitive function QoL-AD	Reduced burden and improved outcomes	Internet-based	Synchronous
18	Scott et al [49] (2018)	Germany	Scoping review	mHealth apps	—	Adoption barriers and caregiver stress	High mobile penetration but limited rural uptake	Smartphone-based	Asynchronous
19	Hengst et al [41] (2023)	UK	Observational	Medication adherence through mHealth	120	Compliance metrics	Improved adherence and socialization	Smartphone-based	Asynchronous
20	Lim et al [50] (2018)	Spain	Observational	Remote monitoring tools	50	Cognitive tracking	Earlier interventions and better outcomes	Internet-based	Asynchronous
21	David et al [42] (2023)	UK	Survey	Specialist video consultation	—	QoL and cognitive scores	Improved outcomes in urban areas	Internet-based	Synchronous
22	Gabb et al [43] (2025)	UK	Observational monitoring study	General telehealth check-ins (wearable devices)	—	Caregiver outcomes	Reduced stress and improved satisfaction	Internet-based	Synchronous
23	Chi and Demiris [44] (2015)	UK	Review	Wearables for health tracking	—	Device accuracy	Effective monitoring and scaling potential	Wearables	Synchronous

[a]Not available.

[b]MMSE: Mini-Mental State Examination.

[c]HVLT-R: Hopkins Verbal Learning Test-Revised.

[d]mHealth: mobile health.

[e]QoL: quality of life.

[f]QoL-AD: quality of life in Alzheimer disease.

XSL•FO
RenderX

Telemedicine interventions range from remote monitoring, mHealth apps, and video consultations, reflecting differences in health care infrastructure across urban and rural Nigeria. The intervention designs were diverse, highlighting the varied nature of telemedicine adoption.

Three studies were RCTs [46,51,52]. Five studies used quantitative methods, including observational and pretest-posttest designs [6,45,47,53,54]. Six studies used qualitative methods focusing on interviews and observational data to explore feasibility, acceptability, and implementation challenges [13,34,39,48,55,56]. Nine studies used mixed methods, combining quantitative (eg, surveys) and qualitative (eg, interviews) data for comprehensive evaluation [9,11,28,33,35,41,43,49,50]

Sample sizes varied, the largest study involving 250 participants [35] and the smallest only 11 participants [37]. Three studies reported sample sizes over 100 participants [35,45,53], with larger studies using quantitative methods and smaller ones exploring qualitative insights and feasibility. Outcome measure varied, caregiver burden was assessed in 6 studies using self-reported questionnaires [6,11,32,35,39,45]. Four studies used standardized scales to measure quality of life among individuals living with dementia, such as the QoL-AD scale [6,45,46,54]. Studies in rural areas often identified internet access and digital literacy as critical secondary outcomes [6,48]. Two studies addressed behavioral and psychological outcomes, highlighting reductions in agitation and caregiver stress [34,39]. The projected telemedicine growth in Nigeria from 2018 to 2024 shows increased use of mHealth apps, video consultations, and remote monitoring tools across urban and rural areas. Studies show mHealth apps are widely used in Nigeria for telemedicine, supporting caregivers in managing care routines, monitoring individuals living with dementia's health [33].

Common barriers include poor internet connectivity [40], high-cost platforms, devices, and mobile data for health care professionals and caregivers [38]. High mobile phone penetration supports mHealth adoption, particularly in urban areas [49]. This study highlights the challenges and benefits of telemedicine for dementia care in Nigeria, revealing gaps in infrastructure and cultural adaptation. Table 2 summarizes the studies, detailing locations, study design, sample sizes, interventions, and outcomes.

Types of Intervention

Studies (n=23) were grouped according to the type of telemedicine intervention, including mHealth apps (n=10), video-based interventions (n=8), and remote monitoring interventions (n=5).

mHealth-Based Telemedicine Interventions

Ten studies assessed mHealth interventions for dementia care [11,32,36,41,45,47,52,53,55,57]. These included apps for communication, medication reminders, and stress management [45,55], as well as health-tracking tools for cognitive decline and behavior [36,47,58]. Smartphone-based systems improved usability, adherence, and caregiver support [11,32,41,52,53], though barriers such as cost, internet access, and digital literacy persist [11,52].

Video Consultations

Eight studies [9,13,34,35,42,46,53,58] evaluated video consultations linking individuals with dementia and their caregivers to health professionals . These interventions improved specialist care coordination, proved most effective for monitoring and assessments in urban areas, and reduced caregiver travel burden while enhancing access to specialists, although their effectiveness in rural settings was limited by internet connectivity challenges.

Remote Monitoring Tools

Five studies [6,34,39,41,50] investigated remote monitoring tools, including wearable sensors and home-based systems, for assessing cognitive function, daily activity, and sleep patterns. While these technologies demonstrated potential for early detection and improved care management, their adoption was hindered by costs and limited digital literacy, particularly in rural areas.

Caregiver Support Apps

Four studies evaluated apps supporting caregivers [11,36,53,55]. These tools improved communication and reminders, reducing stress [55], enhancing well-being [53], and supporting telehealth and national strategies [11,36].

Remote Health Tracking Apps

Three studies [36,55,58] addressed mHealth apps monitoring cognitive decline and quality of life. Yi et al [36] evaluated medication adherence and cognitive tracking, while Zou et al [57] showed reduced hospital visits, Hengst et al [41] noted improved medication adherence and socialization, while Zhu et al [47] highlighted effective symptom tracking and reduced emergency visits.

Video-Based Interventions

Teleconsultations for Specialist Care

Three studies [7,13,37] found video consultation improved care access without in-person visits. Ruggiero et al [59] reported improved quality of life and cognitive scores, especially in urban areas. Chi and Demiris [44] noted reduced stress and better patient outcomes through checkups.

General Remote Video Visits

Two studies examined routine video consultations for dementia care, which addressed caregiver concerns but produced mixed results [1,46]. While Mulili et al. [1] highlighted that poor connectivity limited adoption in rural areas, Page et al [46] reported reduced caregiver burden through their regular use.

Remote Monitoring Tools

Four studies [11,13,35,53] examined home-based and wearable sensors. Cote et al [39] evaluated wearable sensors tracking daily activities, enabling early interventions. Ojeahere et al [34] highlighted monitoring devices tracking behavioral changes and improving care.

Home-Based Systems

Home systems monitored daily activities, cognitive health, and behaviors, Cote et al [39] noted early detection of health

XSL•FO
RenderX

declines, while Ojeahere et al [34] reported reduced caregiver workload, particularly useful in rural areas.

Wearable Sensors

Wearable sensors improved health tracking accuracy and interventions. Chi and Demiris [44] noted effective monitoring, while Louis et al [6] emphasized cost challenges but potential for scaling in resource-poor settings.

Internet Access

Rezigalla [54] and Ojeahere et al [34] discussed the challenges posed by limited internet connectivity in rural areas, which hindered the implementation of telemedicine tools. These studies suggested infrastructure improvements as a critical step to expanding telemedicine access in such regions.

Digital Literacy

Digital literacy was a major barrier noted by Gately et al [32] and Oyinlola et al [11]. Many caregivers and health care providers lacked the skills to use telemedicine platforms effectively, reducing their uptake. The studies recommended comprehensive training to improve digital competency and telemedicine effectiveness.

Challenges and Feasibility

Several studies highlighted the challenges that remain despite the benefits of remote monitoring. Louis et al [6] reported improvements in urban care management, but clinician readiness, poor internet, device costs, and legal concerns hindered rural adoption. Similar barriers were noted in Germany by Mulili et al [1], where lack of training, financial challenges, and inadequate health care infrastructure limited remote monitoring effectiveness. Travers et al [53] noted that low-resource settings faced technical literacy issues, restricting the scalability of interventions. Other interventions such as telepsychiatry and online support networks were explored [60,61]. Aderinto et al [61] studied telepsychiatry in rural Nigeria, showing potential to address mental health disparities in dementia care. Oyinlola et al [11] evaluated online support groups, noting reduced stress and caregiver burden through resource sharing and social support.

Telemedicine Platforms and Legal Frameworks

The scope of telemedicine platforms and regulatory frameworks affecting dementia care in low-resource settings was examined [14]. In rural areas, regulatory barriers to adoption were identified, indicating the need for supportive legal frameworks. Hamilton and Finley [56] reviewed platforms, noting that effective ones improved patient satisfaction and outcomes but were hindered by cost and accessibility issues.

Gaps in Research

The review identified several gaps in the current research on telemedicine for dementia care in Nigeria. First, lack of large-scale studies: larger studies are needed to assess the long-term impact of telemedicine on dementia care; most studies are small-scale [11]. Second, minimal focus on cultural adaptation: limited research addresses adapting telemedicine tools to meet the cultural needs of caregivers and health care professionals [62]. Third, absence of RCTs: few RCTs have been conducted to strengthen evidence for policy and compare telemedicine with traditional care in Nigeria [63].

Discussion

Summary of the Findings

This scoping review examined telemedicine interventions for dementia care in Nigeria, focusing on video consultations, mHealth, and remote monitoring tools. Video consultations effectively improved specialist access and reduced caregiver burden [11]. mHealth apps are the most used intervention due to high mobile phone penetration [63], while remote monitoring tools face challenges in rural areas due to poor internet connectivity and digital literacy gaps [3]. Key barriers include infrastructural limitations, lack of a cohesive strategy, and cultural barriers [4]. Despite challenges, telemedicine shows potential to enhance patient outcomes, reduce caregiver burden, and improve health care access [56]. The findings align with recent studies demonstrating increased adoption of digital health in low-resource settings [26,37,52]. In addition, successful implementation models, such as remote patient monitoring in dementia care, have been demonstrated in previous studies that reported the successful implementation of remote patient monitoring for dementia care [42,57].

Detailed Discussion of Findings

Video consultations address the shortage of dementia specialists by reducing caregiver travel stress and providing timely health care access [23]. However, poor internet access and inconsistent electricity supply hinder implementation in rural areas, such as other sub-Saharan nations facing infrastructural challenges [5,53]. The extensive use of mHealth apps helps caregivers manage routines and monitor health [4]. High mobile phone penetration supports mHealth adoption, particularly in urban areas, but gaps in digital literacy limit rural use. This issue is consistent with other low- and middle-income countries facing similar telemedicine challenges [64].

Remote monitoring tools, such as wearable sensors and home-based systems, support continuous health monitoring and early detection of cognitive decline, reducing hospital visits [58]. However, affordability, digital literacy, and infrastructure challenges limit their use in Nigeria, reflecting similar issues reported globally [45,58].

Cultural and linguistic diversity also impacts telemedicine adoption. Many older adults and caregivers in rural areas are not fluent in English, the primary language of most platforms [65]. Effective adoption requires multilingual support and culturally appropriate frameworks [66]. This aligns with findings from other multicultural regions emphasizing culturally tailored interventions [65].

Limitations

This review has limitations. The methodological quality of studies varies, with many relying on qualitative data, limiting applicability across Nigeria [18]. The review focused on published literature, overlooking relevant gray literature [52]. Few long-term studies evaluate the lasting impact of telemedicine interventions on dementia care outcomes in

XSL•FO
RenderX

Nigeria, with long-term effects still unknown, despite short-term advantages [13]. Lastly, issues with internet connectivity and digital literacy are commonly noted, but strategies to address them were not comprehensively explored [41].

Conclusions

This review highlights the urgent need for a comprehensive telemedicine policy framework integrating technological solutions within existing dementia care systems. Addressing infrastructural limitations, encouraging caregiver training, and developing culturally adapted tools are essential for broader acceptance. Caregiver-centered solutions can support a sustainable model for dementia care in Nigeria. Policymakers should focus on long-term evaluations, RCTs, and scalable models to bridge the digital divide. These strategies will enhance the quality of life for individuals with dementia and establish Nigeria as a leader in innovative health solutions for resource-limited settings.

Acknowledgments

The authors acknowledge the Department of Computing and Informatics, Bournemouth University, United Kingdom. No generative AI tools were used in the writing of this manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

Conceptualization: AA, HD, FA
Formal analysis: AA
Investigation (literature review, data charting): AA, MH
Methodology: AA, HD
Supervision: HD, FA, MH
Writing–Original Draft: AA
Writing–Review & Editing: AA, HD, FA, MH
Visualization (tables, figures, PRISMA flowchart): AA, FA

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed search strategy, study selection process, and full list of included and excluded studies with reasons.
[DOCX File, 17 KB - i-jmr_v14i1e75168_app1.docx]

Checklist 1
Completed PRISMA-ScR 2020 checklist.
[DOCX File, 86 KB - i-jmr_v14i1e75168_app2.docx]

References

1. Muili AO, Mustapha MJ, Offor MC, Oladipo HJ. Emerging roles of telemedicine in dementia treatment and care. Dement Neuropsychol 2023;17:e20220066. [doi: 10.1590/1980-5764-DN-2022-0066] [Medline: 37261258]
2. Mohamed WM. Dementia in AfrAbia: a bibliometric analysis. In: Essential Guide to Neurodegenerative Disorders: Academic Press; 2024. [doi: 10.1016/B978-0-443-15702-8.00019-1]
3. Wakawa IA, Musami UB, Kwairanga SH. Management of dementia in a resource-constrained sub-Saharan African setting: outcome of a retrospective survey of clinical practice in the only neuropsychiatric facility in northeastern Nigeria. medRxiv. Preprint posted online on Sep 10, 2024. [doi: 10.1101/2024.09.09.24313311]
4. Aboye GT, Vande Walle M, Simegn GL, Aerts JM. Current evidence on the use of mHealth approaches in sub-Saharan Africa: a scoping review. Health Policy Technol 2023 Dec;12(4):100806. [doi: 10.1016/j.hlpt.2023.100806]
5. Jones-Esan L, Somasiri N, Lorne K. Enhancing healthcare delivery through digital health interventions. Research Square. Preprint posted online on Oct 3, 2024. [doi: 10.21203/rs.3.rs-5189203/v1]
6. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. Neuro Oncol 2021 Aug 2;23(8):1231-1251. [doi: 10.1093/neuonc/noab106] [Medline: 34185076]
7. Ezeamii VC, Okobi OE, Wambai-Sani H, et al. Revolutionizing healthcare: how telemedicine is improving patient outcomes and expanding access to care. Cureus 2024 Jul;16(7):e63881. [doi: 10.7759/cureus.63881] [Medline: 39099901]

XSL•FO
RenderX

8. Ozili PK. COVID-19 pandemic and economic crisis: the Nigerian experience and structural causes. JEAS 2021 Oct 22;37(4):401-418. [doi: 10.1108/JEAS-05-2020-0074]
9. Ibrahim AM, Alenezi IN, Mahfouz AKH, et al. Examining patient safety protocols amidst the rise of digital health and telemedicine: nurses' perspectives. BMC Nurs 2024 Dec 19;23(1):931. [doi: 10.1186/s12912-024-02591-8] [Medline: 39702255]
10. Bossen AL, Kim H, Williams KN, Steinhoff AE, Strieker M. Emerging roles for telemedicine and smart technologies in dementia care. Smart Homecare Technol Telehealth 2015;3:49-57. [doi: 10.2147/SHTT.S59500] [Medline: 26636049]
11. Oyinlola O, Mahmoud K, Adeoti AB, Abiodun AA. COVID-19: Experiences of social workers supporting older adults with dementia in Nigeria. Inquiry 2024;61:469580241239844. [doi: 10.1177/00469580241239844] [Medline: 38500245]
12. Hwang TJ. Telehealth & mental healthcare of older adults. IPA bulletin. International Psychogeriatric Association. 2020. URL: https://www.ipa-online.org/UserFiles/file/IPABulletin2020.Telehealth.SpecialIssue_v1.pdf [accessed 2025-05-27]
13. Angelopoulou E, Papachristou N, Bougea A, et al. How telemedicine can improve the quality of care for patients with Alzheimer's disease and related dementias? A narrative review. Med Bogota Colomb 2022;58(12):1705. [doi: 10.3390/medicina58121705]
14. Anthony Jnr B. Implications of telehealth and digital care solutions during COVID-19 pandemic: a qualitative literature review. Inform Health Soc Care 2021 Mar 2;46(1):68-83. [doi: 10.1080/17538157.2020.1839467] [Medline: 33251894]
15. Uwishema O, Frederiksen KS, Correia IFS, Mahmoud A, Onyeaka H, Dost B. The impact of COVID-19 on patients with neurological disorders and their access to healthcare in Africa: a review of the literature. Brain Behav 2022 Sep;12(9):e2742 [FREE Full text] [doi: 10.1002/brb3.2742] [Medline: 35951730]
16. El-Sherif DM, Abouzid M, Elzarif MT, Ahmed AA, Albakri A, Alshehri MM. Telehealth and artificial intelligence insights into healthcare during the COVID-19 pandemic. Healthcare (Basel) 2022 Feb 18;10(2):385. [doi: 10.3390/healthcare10020385] [Medline: 35206998]
17. Alade M, Sanusi B. Alternative to health consultations or safe spaces? Pregnant women's health information-seeking behaviour online during the COVID-19 pandemic in Nigeria. GKMC 2024. [doi: 10.1108/GKMC-11-2023-0440]
18. Haimi M. The tragic paradoxical effect of telemedicine on healthcare disparities- a time for redemption: a narrative review. BMC Med Inform Decis Mak 2023 May 16;23(1):95. [doi: 10.1186/s12911-023-02194-4] [Medline: 37193960]
19. Ogunniran IA, Owolabi RO, Musa AA, Magaji AA, Aliyu IU. Role of telemedicine in increasing healthcare access in Nigeria: challenges and effective adoption. Compr Health Biomed Stud 2023;2(2):e153991. [doi: 10.5812/chbs-153991]
20. Fabricatore C, Radovic D, Lopez X, Grasso-Cladera A, Salas CE. When technology cares for people with dementia: A critical review using neuropsychological rehabilitation as a conceptual framework. Neuropsychol Rehabil 2020 Sep;30(8):1558-1597. [doi: 10.1080/09602011.2019.1589532] [Medline: 30896282]
21. Haleem A, Javaid M, Singh RP, Suman R. Telemedicine for healthcare: capabilities, features, barriers, and applications. Sens Int 2021;2:100117. [doi: 10.1016/j.sintl.2021.100117] [Medline: 34806053]
22. Sorrentino M, Fiorilla C, Mercogliano M, et al. Technological interventions in European dementia care: a systematic review of acceptance and attitudes among people living with dementia, caregivers, and healthcare workers. Front Neurol 2024;15:1474336. [doi: 10.3389/fneur.2024.1474336] [Medline: 39416661]
23. Payne RE, Clarke A. How and why are video consultations used in urgent primary care settings in the UK? A focus group study. BJGP Open 2023 Sep;7(3):BJGPO.2023.0025. [doi: 10.3399/BJGPO.2023.0025] [Medline: 37068795]
24. Martis CS, Bhandary RP, Chandrababu R, et al. Caring burden and quality of life among the caregivers of people living with dementia - a cross-sectional study in Udupi district of Karnataka. Home Health Care Serv Q 2024;43(3):191-204. [doi: 10.1080/01621424.2023.2301417] [Medline: 38190733]
25. Chitungo I, Mhango M, Mbunge E, Dzobo M, Musuka G, Dzinamarira T. Utility of telemedicine in sub-Saharan Africa during the COVID-19 pandemic. A rapid review. Hum Behav Emerg Technol 2021 Dec;3(5):843-853. [doi: 10.1002/hbe2.297] [Medline: 34901772]
26. Oluokun EO, Adedoyin FF, Dogan H, Jiang N. Digital interventions for managing medication and health care service delivery in West Africa: systematic review. J Med Internet Res 2024 Oct 9;26:e44294. [doi: 10.2196/44294] [Medline: 39383531]
27. Asare AK, Bannor R, Yawson RM, Bawole JN. Healthcare system innovation - the case of adoption of telemedicine in Ghana. IJBSR 2023;17(4):407-441. [doi: 10.1504/IJBSR.2023.131722]
28. Durmuş A. The influence of digital literacy on mHealth app usability: the mediating role of patient expertise. Digit Health 2024;10:20552076241299061. [doi: 10.1177/20552076241299061] [Medline: 39600388]
29. Adedeji T, Fraser H, Scott P. Implementing electronic health records in primary care using the theory of change: Nigerian case study. JMIR Med Inform 2022 Aug 11;10(8):e33491. [doi: 10.2196/33491] [Medline: 35969461]
30. Bello O, Ajao AO. Digital literacy and skills development in Nigeria: policies, barriers and recommendations. AJAIAS 2024;5(2) [FREE Full text]
31. Dodoo JE, Al-Samarraie H, Alzahrani AI. Telemedicine use in Sub-Saharan Africa: barriers and policy recommendations for COVID-19 and beyond. Int J Med Inform 2021 Jul;151:104467. [doi: 10.1016/j.ijmedinf.2021.104467] [Medline: 33915421]

XSL•FO
RenderX

32. Gately ME, Trudeau SA, Moo LR. In-home video telehealth for dementia management: implications for rehabilitation. Curr Geriatr Rep 2019 Sep 1;8(3):239-249. [doi: 10.1007/s13670-019-00297-3] [Medline: 32015957]
33. Adenuga KI, Farnborough College of Technology. Telemedicine system: service adoption and implementation issues in Nigeria. IJST 2020 Mar 25;13(12):1321-1327. [doi: 10.17485/IJST/v13i12.180]
34. Ojeahere MI, de Filippis R, Ransing R, et al. Management of psychiatric conditions and delirium during the COVID-19 pandemic across continents: lessons learned and recommendations. Brain Behav Immun Health 2020 Dec;9:100147. [doi: 10.1016/j.bbih.2020.100147] [Medline: 32984854]
35. Gaugler JE, Zmora R, Mitchell LL, et al. Remote activity monitoring for family caregivers of persons living with dementia: a mixed methods, randomized controlled evaluation. BMC Geriatr 2021 Dec 18;21(1):715. [doi: 10.1186/s12877-021-02634-8] [Medline: 34922475]
36. Yi JS, Pittman CA, Price CL, Nieman CL, Oh ES. Telemedicine and dementia care: a systematic review of barriers and facilitators. J Am Med Dir Assoc 2021 Jul;22(7):1396-1402. [doi: 10.1016/j.jamda.2021.03.015] [Medline: 33887231]
37. 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]
38. Arora S, Huda RK, Verma S, Khetan M, Sangwan RK. Challenges, barriers, and facilitators in telemedicine implementation in India: a scoping review. Cureus 2024 Aug;16(8):e67388. [doi: 10.7759/cureus.67388] [Medline: 39310647]
39. Cote AC, Phelps RJ, Kabiri NS, Bhangu JS, Thomas KK. Evaluation of wearable technology in dementia: a systematic review and meta-analysis. Front Med (Lausanne) 2020;7:501104. [doi: 10.3389/fmed.2020.501104] [Medline: 33505979]
40. Deniz-Garcia A, Fabelo H, Rodriguez-Almeida AJ, et al. Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. J Med Internet Res 2023 May 4;25:e44030. [doi: 10.2196/44030] [Medline: 37140973]
41. Hengst TM, Lechner L, Dohmen D, Bolman CA. The facilitators and barriers of mHealth adoption and use among people with a low socio-economic position: a scoping review. Digit Health 2023;9:20552076231198702. [doi: 10.1177/20552076231198702] [Medline: 37691766]
42. David MCB, Kolanko M, Del Giovane M, et al. Remote monitoring of physiology in people living with dementia: an observational cohort study. JMIR Aging 2023 Mar 9;6:e43777. [doi: 10.2196/43777] [Medline: 36892931]
43. Gabb VG, Blackman J, Morrison H, et al. Longitudinal remote sleep and cognitive research in older adults with mild cognitive impairment and dementia: prospective feasibility cohort study. JMIR Aging 2025 May 28;8:e72824. [doi: 10.2196/72824] [Medline: 40435500]
44. Chi NC, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. J Telemed Telecare 2015 Jan;21(1):37-44. [doi: 10.1177/1357633X14562734] [Medline: 25475220]
45. Strini V, Prendin A, Cerrone V, et al. Scale of assessment of caregiver care burden of people with dementia: a systematic review of literature. Transl Med UniSa 2023;25(2):38-49. [doi: 10.37825/2239-9747.1044] [Medline: 38343418]
46. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89. [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
47. Zhu L, Xing Y, Jia H, Xu W, Wang X, Ding Y. Effects of telehealth interventions on the caregiver burden and mental health for caregivers of people with dementia: a systematic review and meta-analysis. Aging Ment Health 2024 Nov;28(11):1427-1439. [doi: 10.1080/13607863.2024.2371480] [Medline: 38946249]
48. Ajiboye TO, Oyewo OA, Onwudiwe DC. Simultaneous removal of organics and heavy metals from industrial wastewater: a review. Chemosphere 2021 Jan;262:128379. [doi: 10.1016/j.chemosphere.2020.128379] [Medline: 33182079]
49. 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]
50. Lim RH, Sharmeen T. Medicines management issues in dementia and coping strategies used by people living with dementia and family carers: a systematic review. Int J Geriatr Psychiatry 2018 Dec;33(12):1562-1581. [doi: 10.1002/gps.4985] [Medline: 30270451]
51. Lai FY, Yan EH, Yu KY, Tsui WS, Chan DH, Yee BK. The protective impact of telemedicine on persons with dementia and their caregivers during the COVID-19 pandemic. Am J Geriatr Psychiatry 2020 Nov;28(11):1175-1184. [doi: 10.1016/j.jagp.2020.07.019]
52. Kruse C, Heinemann K. Facilitators and barriers to the adoption of telemedicine during the first year of COVID-19: systematic review. J Med Internet Res 2022 Jan 4;24(1):e31752. [doi: 10.2196/31752] [Medline: 34854815]
53. Travers JL, Wittenberg GF, Gifford DR, Reddy A, McLaughlin MM, Baier RR. Providers' perspectives on high-quality dementia care in long-term care. J Am Med Dir Assoc 2022 Dec;23(12):2030. [doi: 10.1016/j.jamda.2022.07.025] [Medline: 36058295]
54. Rezigalla AA. Observational study designs: synopsis for selecting an appropriate study design. Cureus 2020 Jan 17;12(1):e6692. [doi: 10.7759/cureus.6692] [Medline: 31988824]
55. Levoy K, Rivera E, McHugh M, Hanlon A, Hirschman KB, Naylor MD. Caregiver engagement enhances outcomes among randomized control trials of transitional care interventions: a systematic review and meta-analysis. Med Care 2022 Jul 1;60(7):519-529. [doi: 10.1097/MLR.0000000000001728] [Medline: 35679175]

XSL•FO
RenderX

56. Hamilton AB, Finley EP. Qualitative methods in implementation research: an introduction. Psychiatry Res 2019 Oct;280:112516. [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
57. Zou N, Xie B, He D, Hilsabeck R, Aguirre A. mHealth Apps for dementia caregivers: systematic examination of mobile apps. JMIR Aging 2024 Nov 20;7:e58517. [doi: 10.2196/58517] [Medline: 39621936]
58. Ärlebrant L, Dubois H, Creutzfeldt J, Edin-Liljegren A. Emergency care via video consultation: interviews on patient experiences from rural community hospitals in northern Sweden. Int J Emerg Med 2024 Sep 3;17(1):109. [doi: 10.1186/s12245-024-00703-4] [Medline: 39227787]
59. Ruggiero F, Zirone E, Molisso MT, et al. Telemedicine for cognitive impairment: a telephone survey of patients' experiences with neurological video consultation. Neurol Sci 2023 Nov;44(11):3885-3894. [doi: 10.1007/s10072-023-06903-9] [Medline: 37365397]
60. Al Shamsi H, Almutairi AG, Al Mashrafi S, Al Kalbani T. Implications of language barriers for healthcare: a systematic review. Oman Med J 2020 Mar;35(2):e122. [doi: 10.5001/omj.2020.40] [Medline: 32411417]
61. Aderinto N, Olatunji G, Kokori E, et al. Telepsychiatry in Africa: overcoming barriers to access and care. Discov Ment Health 2025 Apr 29;5(1):64. [doi: 10.1007/s44192-025-00197-3] [Medline: 40299230]
62. Akinyemi RO, Yaria J, Ojagbemi A, et al. Dementia in Africa: current evidence, knowledge gaps, and future directions. Alzheimers Dement 2022 Apr;18(4):790-809. [doi: 10.1002/alz.12432] [Medline: 34569714]
63. Kruse CS, Molina-Nava A, Kapoor Y, Anerobi C, Maddukuri H. Analyzing the effect of telemedicine on domains of quality through facilitators and barriers to adoption: systematic review. J Med Internet Res 2023 Jan 5;25:e43601. [doi: 10.2196/43601] [Medline: 36602844]
64. Ayo-Farai O, Ogundairo O, Maduka CP, Okongwu CC, Babarinde AO, Sodamade OT. Telemedicine in health care: a review of progress and challenges in Africa. Matrix Science Pharma 2023;7(4):124-132. [doi: 10.4103/mtsp.mtsp_24_23]
65. Nielsen TR, Nielsen DS, Waldemar G. Feasibility of a culturally tailored dementia information program for minority ethnic communities in Denmark. Int J Geriatr Psychiatry 2022 Jan;37(1). [doi: 10.1002/gps.5656] [Medline: 34762345]
66. Jegede OO. Evaluating the potential of a unified Hausa-Igbo-Yoruba language to ease language related social and political conflicts in Nigeria. J Univers Lang 2024 Sep;25(2):51-92. [doi: 10.22425/jul.2024.25.2.51]

Abbreviations

mHealth: mobile health
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by TDA Cardoso; submitted 28.03.25; peer-reviewed by AO Agbeyangi, A Adekola, SB Guo; revised version received 15.07.25; accepted 17.07.25; published 27.10.25.

Please cite as:
Adedeji A, Dogan H, Adedoyin F, Heward M
Adoption of Telemedicine for Dementia Care in Nigeria: Scoping Review
Interact J Med Res 2025;14:e75168
URL: https://www.i-jmr.org/2025/1/e75168
doi:10.2196/75168

© Abiodun Adedeji, Huseyin Dogan, Festus Adedoyin, Michelle Heward. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.10.2025. 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

Obesity Prevention and Reduction in China Using the Social Media Platform WeChat: Scoping Review

Yinuo Wang[1], MM; Xuxiu Zhuang[2], MD; Samantha Sundermeir[3], PhD; Joel Gittelsohn[3], PhD

[1]Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[2]Institute of Health Sciences, China Medical University, Shenyang, China

[3]Center for Human Nutrition, Department of International Health, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe Street, Baltimore, MD, United States

Corresponding Author:
Joel Gittelsohn, PhD
Center for Human Nutrition, Department of International Health, Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe Street, Baltimore, MD, United States

Abstract

Background: Digital interventions for obesity have demonstrated efficacy in obesity prevention and management. The emergence of smartphones and ubiquitous apps such as WeChat represents potential modality to enhance the reach, sustainability, and cost-effectiveness of such interventions. By the end of the first quarter of 2024, WeChat had approximately 1.36 billion monthly active users, accounting for 96.5% of China's population. The use of this platform for obesity interventions has been validated in multiple Chinese trials, most published in Chinese language journals.

Objective: We aim to synthesize the existing evidence on obesity interventions delivered through WeChat to generate implications for future intervention design and development, thereby reaching an international audience.

Methods: We conducted a scoping review of PubMed and China National Knowledge Infrastructure using search terms including "WeChat," "obesity," "weight," "BMI," "waist circumference," "hip circumference," "waist-to-hip ratio," "body fat," "skin fold thickness," and these Chinese equivalents "weixin," "feipang," "tizhong," "tizhongzhishu," "yaowei," "tunwei," "yaotunbi," "tizhi," and "pizhehoudu." We included only original research studies, theses, or dissertations with measurable outcomes that used WeChat functions as intervention strategies. Study quality was assessed using the National Institutes of Health Quality Assessment Tool, with specific tools selected based on study design. Descriptive statistics were applied, with categorical variables summarized as frequencies and percentages (n, %) to report study distribution.

Results: Our scoping review based on PubMed and China National Knowledge Infrastructure identified 665 initial records, among which 43 studies met eligibility criteria and were included for data extraction to characterize intervention details. Results indicated effectiveness in 86.0% (37/43) of studies, with WeChat-assisted obesity interventions achieving significant short- and long-term weight loss measured by objective outcomes (body weight, BMI, waist circumference, hip circumference, waist-to-hip ratio, and body fat percentage). However, formative research informing intervention design was insufficient. Common methodological limitations included lack of randomization and blinding (42/43, 97.7%) and unreported intervention compliance metrics (39/43, 92.0%). Functionally, interventions primarily used "WeChat group" and "Official Account"—public accounts that provide health education, diet or physical activity logging, and other features.

Conclusions: Overall, WeChat represents a promising platform for obesity interventions; however, current apps fail to leverage its full features (eg, online payment and live streaming). Key limitations include methodological heterogeneity and cultural specificity, which were addressed through narrative synthesis stratified by study types. Future research should incorporate the formative phase and use more rigorous methodologies such as randomized controlled trials to optimize intervention design and delivery via this modality.

(Interact J Med Res 2025;14:e65538) doi:10.2196/65538

KEYWORDS

obesity; digital interventions; mHealth; WeChat; scoping review; prevention; public health; Chinese; PRISMA; Preferred Reporting Items for Systematic Reviews and Meta-Analyses

XSL•FO
RenderX

Introduction

Obesity is a global health crisis affecting low- and middle-income as well as high-income countries [1,2]. According to the 2023 Chinese Nutrition and Chronic Diseases Survey, more than 50% of Chinese adults were overweight or obese [3], with prevalence rates of nearly 20% among adolescents and 10% in children younger than 6 years [3]. Obesity is strongly associated with noncommunicable chronic diseases (NCDs), particularly fatty liver, diabetes, hypertension, and dyslipidemia [4]. Furthermore, NCDs accounted for 90.1% (9.6 million) of deaths and 84.9% (324.6 million) of disability-adjusted life years in China in 2019 [5]. Wang et al [6] project that the prevalence of overweight or obesity in Chinese adults will reach 65.3% (790.0 million) by 2030, potentially leading to a 22.0% (more than CNY 400.0 billion [more than US $60 billion]) increase in total medical costs. In response, the Healthy China Action Plan 2019‐2030 established 15 key strategies, including 11 focused on NCDs intervention [7]. Thus, implementing evidence-based obesity interventions is critical for preventing future NCD-related mortality.

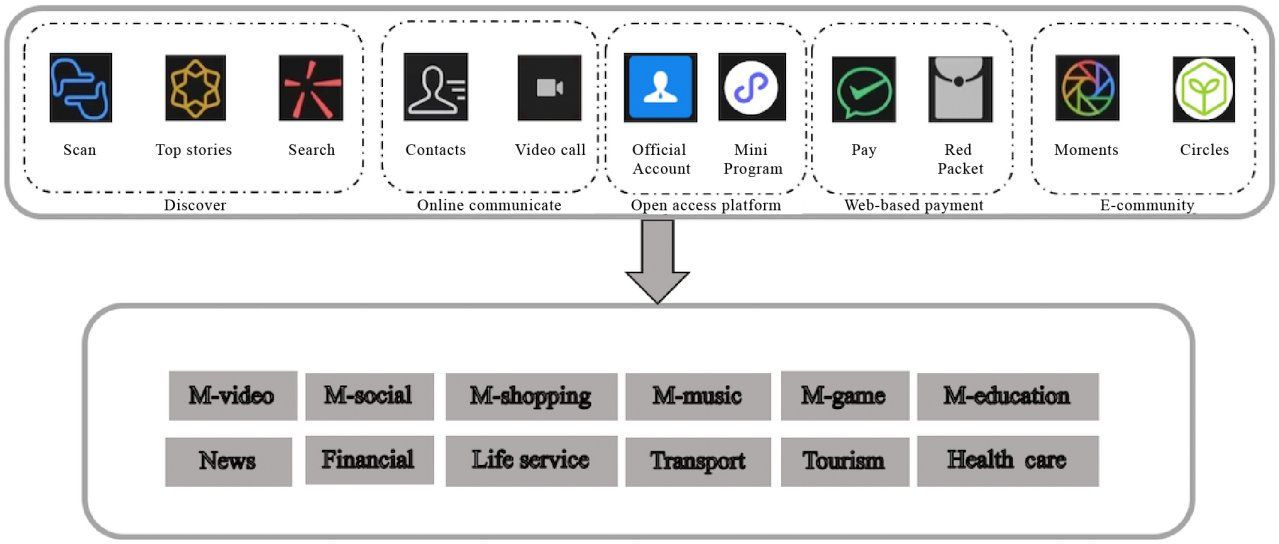
Although nearly 90% of countries globally have committed to addressing obesity [8], the efficacy of interventions for sustained weight control remains limited [9,10]. Social media platforms are increasingly used in obesity interventions, particularly for health education delivery [11,12]. Digital health apps can expand geographical reach and overcome barriers to in-person care [13], as demonstrated during the COVID-19 pandemic when obesity interventions rapidly shifted toward mobile health (mHealth) solutions [14]. In the United States, of more than 100,000 available health-related apps, more than 28,000 target weight management [15]. By late 2023, China had 1.09 billion web users (77.5% penetration) [16], with web-based health care users reaching 413.9 million—a 14.2% annual increase [16]. By reducing medical expenses, improving adherence, and promoting communication between researchers and participants [17,18], mHealth enables cost-effective and individualized interventions [19,20]. According to a systematic review about weight management mobile technology, there are a great number of apps designed to facilitate obesity interventions [21]. For example, Oviva's NHS Digital–approved app is a convenient way to track weight, keep a food diary, and deliver dietary guidance [22-25]; Kurbo is an app designed to help teenagers and children aged 8-17 years to lose weight, using the Traffic Light System designed by Stanford University [26-28]; W8Loss2Go (Weight Loss to Go), implementing a 3-phase approach (identifying problem foods, minimizing eating between meals, and gradually reducing portion sizes), is effective for weight loss [29]; Weight Watchers Online is a weight loss app that focuses on food, physical activity (PA), sleep, and mindset [30,31]; and BeHealthy provides education on nutrients, sleep, and yoga [32]. While designed for weight management guidance and monitoring, few apps incorporate health care professional input or rigorous validation [33].

WeChat, a dominant media platform with multiple functions launched by the Tencent Company in China on January 21, 2011, reached 1.359 billion monthly active users in Q1 2024—a 3% increase from March 2023 [34]. Monthly active user, a key metric reflecting monthly engagement with WeChat services, demonstrates its pervasive adoption. Although initially the primary function of WeChat was chatting, with version updates, WeChat has updated more and more features integrating social networking, mobile payment, and enterprise solutions through features such as Official Account for content dissemination, Mini Program for lightweight apps, and WeChat group for real-time group communication (Figure 1). The platform's multifunctional ecosystem facilitates diverse aspects of users' lives (Figure 1), including health education delivery [35] and chronic disease management [36] such as hypertension [37,38] and diabetes [39-42].

Although previous studies have summarized the efficacy of mHealth apps for weight management, none have systematically detailed the functional implementation of WeChat in obesity interventions or established evidence-based frameworks for optimizing these functions in future projects. The goal of this study was to conduct a scoping review of existing obesity intervention studies through the WeChat platform and its potential use in future research. We focused on answering the following questions: (1) What WeChat functions have been used in existing obesity prevention or management intervention studies?; (2) What are the primary methodological limitations in these studies?; (3) How can future WeChat-delivered interventions be optimized?

XSL•FO
RenderX

Figure 1. Functions of the WeChat platform.



Methods

Search Strategy

To identify literature on WeChat-based obesity interventions up to April 14, 2024, we searched PubMed and China National Knowledge Infrastructure (CNKI, the largest academic database, aggregating more than 95% of officially published Chinese scholarly resources), using the following search terms: "WeChat," "obesity," "weight," "BMI," "waist circumference (WC)," "hip circumference (HC)," "waist-to-hip ratio (WHR)," "body fat (BF)," "skin fold thickness," and Chinese equivalent characters: "weixin," "feipang," "tizhong," "tizhongzhishu," "yaowei," "tunwei," "yaotunbi," "tizhi," and "pizhehoudu" (Table S1 in Multimedia Appendix 1).

Study Selection

After eliminating duplicates and off-topic studies, all the papers that described WeChat-based obesity interventions were screened against the following inclusion criteria: (1) full-text original or empirical human research published in peer-reviewed journals (including trials with or without control groups, and blinded or nonblinded designs); (2) WeChat as primary intervention delivery platform (eg, dietary or PA guidance, dietary or PA diary logging, etc); (3) Chinese degree theses or dissertations from the CNKI dataset; and (4) objective anthropometric outcomes including body weight, BMI, WC, HC, WHR, body fat percentage (BFP), or skin fold thickness. The exclusion criteria included (1) studies using WeChat solely for participant recruitment or follow-up; (2) self-reported anthropometric parameters; and (3) study protocols, animal studies, reviews, patents, books, or inaccessible texts. This process was done by 2 authors (YNW and XXZ) who independently screened titles and abstracts and conducted full-text reviews to identify the final set of eligible papers for inclusion. Any conflicts during this process were resolved through team consensus discussions. The flow diagram was adapted from the official PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) template with modifications specific to our search strategy [43].

Data Extraction

The following data were extracted from each study: authors, sample size, presence of control groups, WeChat functions used, interventions duration, outcomes measured, and intervention effectiveness. Intervention effectiveness was categorized as (1) "effective" if all predefined weight-related outcomes (eg, BMI, weight, WC, HC, WHC, BFP, etc) showed statistically significant improvements; (2) "partially effective" if only a subset of predefined outcomes demonstrated significant positive changes; and (3) "not" if none of the predefined outcomes demonstrated significant changes. Themes were synthesized from tile or abstract review followed by full-text review using a blinded, independent dual review process by 2 of the authors (YNW and XXZ). Conflicts were resolved through discussion, with a third author (JG) who would arbitrate unresolved conflicts.

Quality Assessment

A detailed quality assessment was conducted using the Quality Assessment Tool for Controlled Intervention Studies and Before-After (Pre-Post) Studies With No Control Group designed by National Institutes of Health (NIH) [44]. The assessment tool of controlled intervention studies is applicable to those with control group and intervention strategies, containing 14 items to evaluate the research. Before-after studies refer to those without control groups, instead comparing outcomes of baseline and the end. The tool for before-after studies consists of 12 items. All scoring criteria were evaluated as "yes," "no," "not applicable (NA)," or "not reported (NR)." For each item, we assigned 1 point for "yes" and "NA," zero points for "no" and "NR." According to the total score, studies were further rated as "good" (10 for controlled intervention studies; 9 for before-after studies), "fair" (5-9 for controlled intervention studies; 5-8 for before-after studies), and "poor" (zero to 4 for both). Two authors (YNW and XXZ) independently completed the quality assessment for each study. Conflicts were resolved via discussion, with a third author (JG) arbitrating unresolved cases.

XSL•FO
RenderX

Statistical Analysis

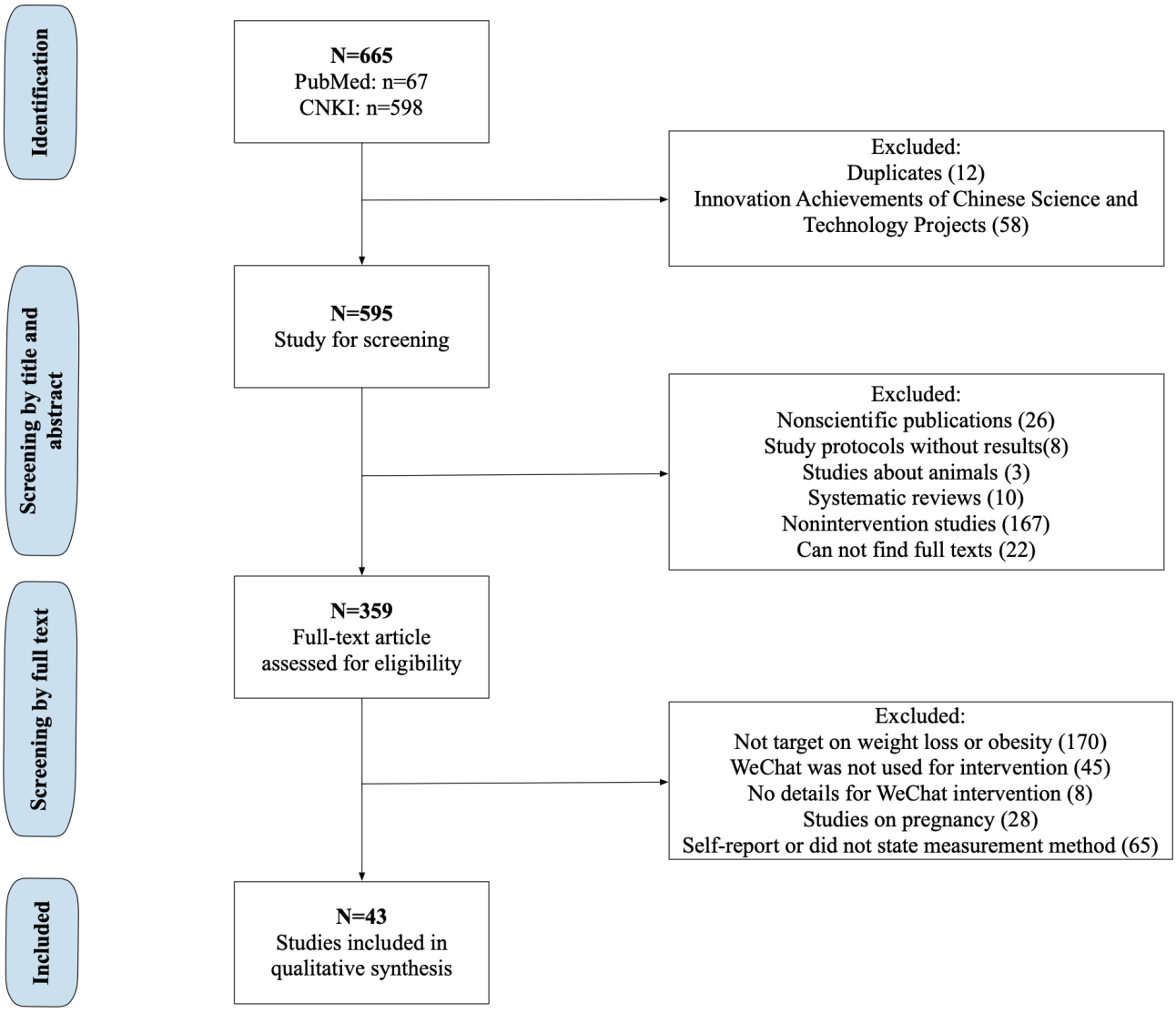
Statistical analyses were performed using GraphPad Prism 8.4.0 (GraphPad). Descriptive statistics were applied, with categorical variables summarized as frequencies and percentages (n, %) to report study distribution across classifications, as well as the quality assessment outcomes of included studies.

Results

Overview

From the initial 665 papers (67 from PubMed and 598 from CNKI), 12 duplicates and 58 Innovation Achievements of Chinese Science and Technology Project designs were excluded. During the first round of title or abstract review, 236 publications were excluded, in which 26 of them were nonscientific publications, 8 study protocols, 3 animal studies, 10 systematic reviews, 167 nonintervention studies, and 22 conference abstracts. Of the 359 studies undergoing full-text review, 315 (87.7%) were excluded due to the following reasons: focus beyond weight loss or obesity (170, 54.0%), WeChat not used for intervention delivery (45, 14.3%), inadequate WeChat intervention strategy details (8, 2.5%), pregnancy-focused studies (28, 8.9%), and self-reported or unverified measures (65, 20.6%). The flow diagram is shown in Figure 2. Eight (18.6%) studies [45-52] were in English with information about study population, control group presence, intervention methods, intervention duration, health outcomes, and intervention effectiveness summarized in Table 1. A total of 35 (35/43, 81.4%) studies [53-87] were in Chinese with study details shown in Table 2. Among the 43 studies finally included, 20 (46.5%) [53,57,59-63,65,70,72,74-76,79,81,83-87] of them were theses or dissertations. All the studies were conducted in mainland China. None of the papers reported using formative research for intervention development. Sample sizes ranged from 20 [64] to 15,310 [46]. Among the 43 papers, 33 (76.7%) [45-47,49-51,56-63,65-67,70-72,74-76,78-87] included a control group, and 26 (60.5%) [45,47,49-51,56-62,65,66,70,72,74,76,78-80,82-86] reported assigning participants randomly. The duration of interventions spanned 4 weeks [51] to 36 months [52].

Figure 2. Flow diagram detailing the study selection process. CNKI: China National Knowledge Infrastructure.



XSL•FO
RenderX

Table . Summary of studies using WeChat for obesity interventions in English (in order of publication date; n=8).

Authors	Sample	Control group	Intervention related to WeChat	Duration	Outcome	Effective[a] or not
Li et al [49]	42 full-time postgraduate students, Shanghai	Yes (randomized)	WeChat: submit PA[b] diaries WeChat group: connect with other participants for social support	1 month	BW[c], BFP[d], WHR[e], and BMI	Partly effective: BFP
Chen et al [47]	102 women from 2 community health centers in Changsha	Yes (randomized)	WeChat: set up fruit or vegetable intake goals and daily steps, educational modules along with tailored tips and messages, and messages to encourage and reinforce positive behavioral changes	6 months	WC[f] and BMI	Not
Han et al [50]	120 overweight and obese patients with T2DM[g] from a hospital in Nanjing	Yes (randomized)	Official Account: information (video teaching, compulsory course, and Monday paper); function (diet form, resistance or aerobic exercise, body measurement, and homework feedback)	12 months	BMI, WC, HC[h], WHR, and BFP	Partly effective: BMI
Xia et al [45]	343 participants with T2DM	Yes (randomized)	WeChat group: post BG test result, completion of exercises, medication, self-management data WeChat voice call: weekly and monthly summaries to participants	6 months	BW and BMI	Partly effective: BW
Yang et al [52]	977 overweight patients from a hospital in Zhejiang	No (before-after)	WeChat group: report anthropometry (eg, body weight), daily dietary intakes, and PA Mini Program: report body weight WeChat: personal feedback on weight management	36 months	BW	Effective
Ling et al [51]	300 inpatients with primary hypertension in a hospital in Xi'an	Yes (randomized)	Official Account: posthypertension-related knowledge WeChat group: posthypertension-related knowledge; distribute information about educational lectures One-to-one chatting: answer questions	1 month	BMI and WC	Not

XSL•FO
RenderX

Authors	Sample	Control group	Intervention related to WeChat	Duration	Outcome	Effective[a] or not
Liu et al [48]	300 overweight or obese participants in Sichuan	No (before-after)	WeChat: postdinner plates and beverage	3 months	BW, BMI, WC, BFP, and WHR	Effective
He et al [46]	15,310 occupational participants in the Shunyi District of Beijing	Yes (those did not follow the Official Account)	Official Account: Health education including intervention of PA and diet	6 months	BW	Partly effective: males

[a]There were favorable changes in the indices after intervention according to their own reports in the study.

[b]PA: physical activity.

[c]BW: body weight.

[d]BFP: body fat percentage.

[e]WHR: waist-to-hip ratio.

[f]WC: waist circumference.

[g]T2DM: type 2 diabetes mellitus.

[h]HC: hip circumference.

XSL•FO
RenderX

Table . Summary of studies using WeChat for obesity interventions in Chinese (in order of publication date; N=35).

Authors	Sample size	Control group	Intervention related to WeChat	Duration	Outcome	Effective[a] or not
Yinbao[b] [84]	1840 from the 51st Regiment of the Third Division in Xinjiang	Yes (randomized)	Official account: post health knowledge	12 months	BW[c], WC[d], HC[e], and BMI	Partly effective: HC and BMI
Lu[b] [83]	62 patients with coronary heart disease in a hospital in Tangshan	Yes (randomized)	Mini Program: record diets WeChat group: supervise diet, share educational material, and healthy diet competitions WeChat live streaming: educational lectures	3 months	BW, BMI, and WC	Partly effective: BW and BMI
Youxuan[b] [74]	295 people at high risk of hypertension from 15 communities in Haikou	Yes (randomized)	WeChat group: reminding to upload diet and PA[f], upload health education, and health guidance	12 months	BMI, WC, and HC	Effective
Jun[b] [86]	60 patients with prediabetes aged 50 - 70 years from a health management center, Guangzhou	Yes (randomized)	WeChat group: share heart rate data, remind to practice, and communication to promote PA	3 months	BW and BFP[g]	Effective
Linjuan[b] [53]	165 patients with T2DM[h] from a hospital in Inner Mongolia	No (before-after)	Mini Program: record glucose, evaluate diet, daily step recording, nutrition recipe, health knowledge, and remind medicine intake	6 months	BMI	Effective
Tianmeng et al [69]	1292 residents with higher blood pressure in Shandong	No (before-after)	WeChat group: post poster or papers about health knowledge; answer daily questions	3 months (100 days)	WC, HC, BW, BMI, and BFP	Effective
Mengqin[b] [76]	480 patients with essential hypertension aged 18–70 years from 5 towns in rural areas of Shanxi	Yes (randomized)	WeChat group: post health knowledge of posters, papers, and videos	3 months	BMI, WC	Not
Wei et al [78]	200 students with normal blood pressure enrolled in 2017 from a university in Xizang	Yes (randomized)	WeChat group: post health education of essays, cartoon, pictures, videos One-to-one chatting: provide consultancy	6 months	BMI	Effective

XSL•FO
RenderX

Authors	Sample size	Control group	Intervention related to WeChat	Duration	Outcome	Effective[a] or not
Ruixue et ali [73]	63 patients with overweight or obesity from a hospital in Chongqing	No (before-after)	WeChat group: post health knowledge, upload personal PA recording, communications among participants, and upload daily diet recording and evaluated by physicians	3 months	BW, BMI, BFP, and WC	Effective
Lu [83]	152 patients with diabetes from 5 communities in Shenzhen	Yes (randomized)	WeChat group: post health knowledge, discussion, and evaluation on nutrition	3 months	BMI and WC	Effective
Xiaoyuan et al [58]	100 patients with phlegm dampness constitution coronary heart disease from a hospital in Nanning	Yes (randomized)	WeRun: record daily steps WeChat group: supervise daily diet and PA	6 months	BW, WC, and BMI	Effective
Zihao et al [54]	54 overweight or obese people from a hospital in Guizhou	No (before-after)	WeChat group: supervise and provide guidance for daily diets and PA; answer questions of participants	3 months	BW, WC, HC, and BFP	Effective
Weiwei[b] [81]	68 older adult patients with chronic diseases with physical inactivity in 2 communities in Bengbu	Yes (participants from Hubin)	WeChat group: post health education of materials, upload weekly PA recording, and health guidance	6 months	BMI and WHR[i]	Effective
Jinhua[b] [59]	107 older adult patients with hypertension from 4 communities in Changsha	Yes (randomized)	WeChat group: share healthy diet and PA videos, upload healthy knowledge about hypertension, and web-based consultancy	6 months	BMI and WC	Partly effective: BMI
Peijun [65]	120 patients with obesity who underwent physical examination from Sichuan University	Yes (randomized)	WeChat group: post health education materials One-to-one chatting: answer questions Official Account: Traditional Chinese Medicine Health Preservation Special Issue	6 months	BW, BMI, BFP, and WHR	Effective
Jielili[b] [75]	48 cases of obesity from Jinglao Community, Hetian	Yes (obesity and normal weight control)	WeChat group: health education; psychological cognition, and weight management	3 months	BW, WC, and HC	Effective

XSL•FO
RenderX

Authors	Sample size	Control group	Intervention related to WeChat	Duration	Outcome	Effective[a] or not
Ziwei[b] [70]	40 young and middle-aged patients with hypertension from 3 community health care centers in Hengyang	Yes (randomized)	WeChat group: upload health education, share BMI and WC change to encourage each other, supervise daily steps, guidance on smoking and drinking, remind on medicine intake, and share music and videos to relax	6 months	BMI and WC	Partly effective: BMI
Yuepeng[b] [61]	40 patients with hyperuricemia in a Tianjin hospital	Yes (randomized)	Official Account: education on diet, PA, and medication knowledge WeChat group: post health education contents, share videos of PA, and answer questions	6 months	BW, WC, HC, BMI, and WHR	Not
Xi et al [55]	47 employees in an affiliated hospital	No (before-after)	WeChat group: share pictures of daily diets and PA; one-to-one dietary guidance by professional nutritionist (including intervention of PA and diet)	1.5 months	BW, BMI, BF, BFP, and WC	Effective
Jianxin et al [71]	130 older adult hypertensive patients from Chinese PLA Hospital No. 305	Yes (not randomized, grouped by first clinic date)	Mini Program: diagnosis and treatment, medication guidance, and data management	6 months	BW, BMI, WC, WHR, and BFP	Partly effective except BW
Li[b] [62]	100 patients with T2DM from community health service center in Chongqing	Yes (randomized)	Official Account: guidance on personal diet and PAs; upload situation of dietary, PA, glucose, and blood pressure; and supervision and health education	12 months	BW, BMI, WC, HC, and BFP	Effective
Yingming[b] [63]	60 patients with overweight or obesity and polycystic ovary syndrome from 2 hospitals in Hainan	Yes (not randomized, grouped by first clinic date)	WeChat group: experience sharing, life behavior guidance, and supervision	3 months	BW, BMI, WC, HC, and WHR	Partly effective except HC
Lijuan[b] [85]	100 young and middle-aged patients with hypertension from all community health service centers in Fuding	Yes (randomized)	WeChat group: health education, consultancy by a doctor, and remind of changing unhealthy timetable	6 months	BW, WC, HC, and BMI	Not
Jiawei[b] [60]	128 patients with T2DM treated or hospitalized in a hospital, Chengdu	Yes (randomly)	Official Account: post intervention strategies, collect information WeChat group: share intervention strategies	3 months	WHR	Effective

XSL•FO
RenderX

Authors	Sample size	Control group	Intervention related to WeChat	Duration	Outcome	Effective[a] or not
Yuexi et al [66]	60 residents with prediabetes from a hospital in Nanning	Yes (randomly)	WeChat group: upload health knowledge, answer questions	24 months	WC and BW	Effective
Jingxia et al [b] [87]	278 employees from 4 worksites in Yang River Delta	Yes (randomly select 2 worksites)	Official Account: post information about obesity, diet and PA, querying energy of food, querying PAs, web-based competition on daily diet, and PAs with rewards	12 months	BW, BMI, WC, HC, and WHR	Partly effective: BW, BMI, and HC
Rongrong [b] [79]	314 individuals with prediabetes diagnosed by a hospital in Shandong	Yes (randomly)	Official Account: health education WeRun: supervise daily steps WeChat group: health lectures and health educations in voice, words, and pictures	7.5 months	BW, BMI, WC, and HC	Effective
Xujuan et al [77]	36 people with obesity from a physical examination center in Taiyuan	No (before-after)	WeChat group: exercise prescription	2 months	WC, BW, BMI, and BFP	Effective
Xiao and Dafang [68]	301 volunteers	No (before-after)	WeChat group: provide guidance on diet, PAs, and psychology	1 month	BW	Effective
Xiaojuan et al [64]	20 overweight patients with prediabetes from a hospital in Wuxi	No (before-after)	WeChat group: upload health knowledge, share daily food and energy intake, and answer questions	6 months	BW, BMI, and WC	Effective
Bei[b] [57]	262 middle-aged and young patients with primary hypertension from a community health service center in Urumqi	Yes (randomly)	WeChat group: upload health education material weekly, answer questions during intervention, and communicate with other participants	6 months	BMI	Not
Huirong et al [80]	223 patients with chronic noninfectious diseases from a hospital in Xinjiang	Yes (Randomly)	Official Account: Public Health Microworld	3 months	BW, BMI, WC, HC, WHR	Effective
Di [b] [72]	40 obesity children aged 10 - 11 from one of primary school in Beijing	Yes (randomly)	WeChat group: share pictures of 3 meals of children; upload web-based education materials (including PA and diet)	2 months	BW, BMI, WC, HC, and BFP	Partly effective: BW and BMI
Jiangping et al [67]	77 college teachers from 2 universities in Harbin	Yes (1 university)	WeChat group: upload health education and share PAs videos	3 months	BMI and WC	Effective

XSL•FO
RenderX

Authors	Sample size	Control group	Intervention related to WeChat	Duration	Outcome	Effective[a] or not
Dongmei et al [56]	86 coronary heart disease patients from a hospital in Shandong	Yes (Randomly)	WeChat group: upload health education, supervise participants to follow personal treatment, and communicate with others	12 months	BMI	Effective

[a]There were favorable changes in the indices after intervention according to their own reports in the study.

[b]Thesis for degree (master's and PhD).

[c]BW: body weight.

[d]WC: waist circumference.

[e]HC: hip circumference.

[f]PA: physical activity.

[g]BFP: body fat percentage.

[h]T2DM: type 2 diabetes mellitus.

[i]WHR: waist-to-hip ratio.

Methods of Intervention

Among the 43 papers ultimately included, 32 (74.4%) [46-48,50,53-57,59,62-64,66-77,80-82,84-87] used a single WeChat function for interventions, while 11 (25.6%) [45,49,51,52,58,60,61,65,78,79,83] used multiple functions of WeChat to implement their intervention studies. “WeChat group” and “Official Account” were 2 functions used the most, with 33 (76.7%) [45,49,51,52,54-61,63-70,72-79,81-83,85,86] and 11 (25.6%) [46,50,51,60-62,65,79,80,84,87] studies using them, respectively (Table 3). WeChat groups primarily facilitated communication between investigators and participants. Official Accounts were mainly used to deliver dietary or PA education and record daily diets and PA. Four (9.3%) studies [52,53,71,83] used Mini Program in their intervention strategies with similar but expanded functions as Official Account. Three (7.0%) studies [51,65,78] used one-to-one chatting functions enabling confidential consultations. In addition, 2 (4.7%) studies [58,79] used WeRun to conduct interventions. Meanwhile, another 2 used voice call [45] and live streaming [83] separately. A total of 4 (9.3%) studies did not clearly report the specific functions of WeChat they used in their intervention (refer to “Intervention” item in Tables 1 and 2 for more details).

Table . WeChat components used in the interventions.

Function	Definition	Values, n (%)
WeChat Group	Containing up to 500 members; all the members can share information and chat in the group; daily attendance of web-based activities	33 (76.7)
Official Account (accounts that developers or merchants apply for on the WeChat official platform)	The platform uses communication and interaction with specific groups of text, pictures, voice, and video, forming a mainstream online and offline WeChat interactive marketing Way	11 (25.6)
Mini Program	Lightweight apps that run within the WeChat environment; smaller than dedicated apps; no need to download	4 (9.3)
One-to-One Chatting	Sending text messages and voice messages; making video calls and voice calls; sharing pictures or WeChat public press paper; transferring money to WeChat friends	3 (7.0)
WeRun	Recording the number of daily steps taken in walking	2 (4.7)
Voice Call	Free communicating method with your WeChat account followers under network environment	1 (2.3)
WeChat Live Streaming	Directly get access to live streams within WeChat without additional download; providing real-time interaction, content sharing, and more	1 (2.3)
Not specified	Did not mention which specific functions were used for intervention	4 (9.3)

XSL•FO
RenderX

Descriptive Statistics on Intervention Efficacy

In summary, 25 out of 43 (58.1%) studies [48,52-56,58,60,62,64-69,73-75,77-82,86] demonstrated significant improvement across all obesity-related measures. Of the studies remaining, 12 (27.9%) [45,46,49,50,59,63,70-72,83,84,87] showed partial effectiveness on the selected measures and 6 (14.0%) [47,51,57,61,78,85] reported no significant effects with a duration period no longer than 6 months. BMI was the most commonly used outcome measure for obesity, in which 27 (77.1%) [48,50,53,55,56,58,59,62-65,67,69-74,77-84,87] out of 35 [45,47-51,53,55-59,61-65,67,69-74,76-85,87] were significantly decreased after intervention. A total of 28 (65.1%) studies [45,46,48,49,52,54,55,58,61-66,68,69,71-73,75,77,79,80,83-87] assessed body weight, in which 23 (82.1%) [45,46,48,52,54,55,58,62-66,68,69,72,73,75,77,79,80,83,86,87] showed weight loss after intervention. A total of 30 (69.8%) studies [47,48,50,51,54,55,58,59,61-64,66,67,69-77,79,80,82-85,87] assessed WC as outcome, in which 18 (60.0%) [48,54,55,58,62-64,66,67,69,71,73-75,77,79,80,82] of them showed a significant reduction after intervention. For HC, a total of 14 (32.6%) studies [50,54,61-63,69,72,74,75,79,80,84,85,87] were considered it as outcome, in which 9 (64.3%) [54,62,69,74,75,79,80,84,87] showed significant decrease. Similarly, a total of 11 (25.6%) studies [48-50,60,61,63,65,71,80,81,87] assessed WHR as outcome, in which 7 (63.6%) [48,60,63,65,71,80,81] of them showed a significant decrease at the end of the study. Of the 14 (32.6%) studies [48-50,54,55,62,65,69,71-73,77,86] assessing BF, only 1 (7.1%) [55] of them used BF and all others used BFP. BF and BFP were shown to be impacted mostly by interventions compared with other measures in these studies because 12 (85.7%) [48,49,54,55,62,65,69,71,73,77,86] out of 14 [48-50,54,55,62,65,69,71-73,77,86] showed that they were significantly decreased (Tables 1 and 2).

Quality Assessment of Study

Using the NIH Study Quality Assessment Tool, we assessed a total of 33 studies [45-47,49-51,56-63,65-67,70-72,74-76,78-87] that had a control group, with 4 (12.1%) [50,70,76,79] rated as good quality, 25 (75.8%) [45,47,49,51,56-63,65-67,71,74,78,81-87] as fair quality, and 4 (12.1%) [46,72,75,80] as poor (See Table S2 in Multimedia Appendix 1). The number of studies meeting each assessment item is shown in Table 4. Twenty-seven (81.8%) [45,47,49-51,56-63,65,66,70,72,74-76,78,79,82-86] of them were described as randomized, in which 16 (59.3%) [50,51,56-60,62,65,66,70,74,76,79,83,85] reported randomization methods that were regarded as adequate (see Table S2 in Multimedia Appendix 1, item 11 and item 22). Only 2 (6.1%) studies [50,76] reported using a double-blinded randomization process, and only 1 (3.0%) study [50] reported that outcome assessors were blinded. Among all the studies with control groups, 6 (18.2%) [49,72,78,80,82,86] did not report demographic information separately for intervention and control groups, and 4 (12.1%) studies [46,66,75,87] reported significant differences in some demographic or other factors related to the outcome being assessed (see Table S2 in Multimedia Appendix 1, item 6). Two (6.1%) studies [45,87] reported an overall dropout rate exceeding 20%, with exact values of 23.1% [45] and 28.4% [87] separately, and 4 (12.1%) studies [65,67,72,75] provided no information about dropout rate (see Table S2 in Multimedia Appendix 1, item 7). In addition, 3 studies reported high intervention adherence [47,72,86], while 3 others reported low adherence [46,61,87]. Besides, intent-to-treat analysis was used in data analysis in 3 studies to avoid bias caused by dropout [47,50,87].

A total of 10 out of 43 (23.3%) studies [48,52-55,64,68,69,73,77] used a before-after design without a control group. According to the rating, all of them [48,52-55,64,68,69,73,77] were rated as fair quality (Table S3 in Multimedia Appendix 1). The number of studies meeting each assessment item is shown in Table 5. All the before-after studies [48,52-55,64,68,69,73,77] stated research questions or objectives clearly and presented specific inclusion and exclusion criteria. Two (20.0%) studies [48,52] included participants who could represent the population who would be eligible, and 1 (10.0%) study [69] enrolled all the eligible participants meeting the entry criteria. Although some of the studies [48,53,54,64,73] provided information about sample size calculations and power to detect statistical differences, they claimed insufficient power as a limitation. All the studies [48,52-55,64,68,69,73,77] clearly and consistently described intervention strategies and used effective and objective outcome measurements. However, none reported whether blinding was used during intervention implementation or data analysis. All studies [48,52-55,64,68,69,73,77] prespecified outcome measures, which were applied clearly and consistently throughout the study process. Three studies reported a loss to follow-up rate exceeding 20%, with rates of 82.4% [52], 24.3% [53], and 55.5% [73], respectively. All studies [48,52-55,64,68,69,73,77] analyzed changes in outcomes of interest before and after the intervention and presented *P* values, but only 3 studies [52,64,68] measured outcomes multiple times before, during, and after the intervention. None of these studies was conducted at the group level; thus, it was not applicable to account for the use of individual-level data in determining group-level effects in the statistical analysis.

XSL•FO
RenderX

Table . Numbers of studies with control group that met each assessment item.

Assessment item	Values, n (%)
Outcomes valid and reliable	33 (100.0)
Outcome prespecified	33 (100.0)
Dropout rate between groups	29 (87.9)
Randomized controlled trial	27 (81.8)
Overall dropout rate	27 (81.8)
Groups similar	23 (69.7)
Randomization adequate	16 (48.5)
Treatment allocation concealed	15 (45.5)
Sample size enough	11 (33.3)
Other interventions avoided	6 (18.2)
High adherence	3 (9.1)
ITT[a] analysis	3 (9.1)
Participants and providers blinded	2 (6.1)
Analyst blinded	1 (3.0)

[a]ITT: intention-to-treat.

Table . Numbers of studies without a control group that met each assessment item.

Assessment item	Value, n (%)
Objective clearly stated	10 (100.0)
Eligibility criteria clear	10 (100.0)
Intervention clear and consistent	10 (100.0)
Outcomes assessment	10 (100.0)
Analyze changes with *P* value	10 (100.0)
Follow-up rate	5 (50.0)
Outcomes measured multiple times	3 (30.0)
Participants representative	2 (20.0)
All eligible participants enrolled	1 (10.0)

Discussion

Principal Findings

This is the first publication to describe intervention studies focused on weight loss or obesity control conducted via WeChat. We identified gaps in the current design and implementation of these digital intervention studies, such as inadequate tracking options and social support, lack of individualized goals, and absence of health care experts [33]. Overall, the results are promising, suggesting that WeChat may be an effective platform for supporting weight loss or obesity management. However, more rigorous, randomized controlled trials are needed to understand the effectiveness of interventions delivered using this platform, and to determine the best implementation strategies. Our research revealed several important findings.

Deficiency of Current Obesity Interventions via WeChat

None of the studies reported using formative research to design the intervention. Formative research is a process of collecting useful data about the targeted population and context before or during intervention implementation, during which behavioral change strategies are developed and refined to be maximally suitable [88]. This process helps with cultural contextualization and optimizes the impact of intervention strategies [89] by generating more effective interventions tailored to specific populations. For example, formative research on diabetes management and support indicated that intervention strategies for type 1 diabetes should be different from those for type 2 [90]. Thus, to better use WeChat in obesity intervention studies, formative research is imperative for developing further strategies to achieve weight management outcomes. Of the 6 studies [47,51,57,61,78,85] that reported no significant changes in obesity-related measures, all had a duration of less than 6 months. Intervention durations of 6-12 months and higher

XSL•FO
RenderX

contact hours were correlated with more substantial changes [91,92]. Thus, future studies should consider extending intervention duration and enhance contact hours.

According to the assessment of studies using the Quality Assessment Tool designed by NIH [44], most studies had insufficient sample sizes. The intervention studies conducted in Chinese populations also had several common limitations. First, although most studies used randomization to assign participants to groups, blinding was rarely used (more than 95% did not report blinding during the intervention process). Similarly, information about participant adherence to interventions was limited, which impeded evaluation of effectiveness. In addition, our review of the inclusion and exclusion criteria found that most studies did not control for confounding effects caused by simultaneous participation in other intervention studies. These limitations might cast doubt on the generalizability of the findings.

Majority WeChat Functions Used in Current Research

WeChat, one of the most widely used social media platforms in China, has many functions that can facilitate obesity interventions [93,94]. In recent years, more intervention studies have used WeChat as a tool in China [41,93-95] because of its convenience and wide reach. The Official Account and WeChat group were the most common features used in the intervention trials included in this review. Both one-to-one chatting and WeChat groups facilitated immediate communication. Although this type of immediate communication might occupy researchers' time, it can also provide quick assistance to participants, potentially reducing loss to follow-up and increasing adherence to interventions, which may lead to positive outcomes. Official Account and Mini Program are user-friendly ways to disseminate intervention-related information, such as health knowledge and PA courses, as well as to record and evaluate personal data on daily diet and PA. While these functions could help researchers deliver health-related information, they do not ensure that participants actually engage with this information.

Generally, WeChat platform shows promise for obesity interventions; however, its utilization for health promotion still needs to be optimized. For example, although researchers can send intervention materials to participants easily via WeChat, there is no guarantee that they will pay attention to the health messages received [75]. Thus, study coordinators should not only use WeChat to deliver health-related information but also try to improve participant engagement through compelling content and strategies to compete for attention with the vast amount of the information disseminated through digital platforms [75].

Future Directions—Use of WeChat in Future Obesity Intervention Research

Online Payment

To date, no study has incorporated the online payment function of WeChat into existing obesity interventions. A potential strategy could involve offering time-limited discounts on healthy foods (such as vegetables and fruits) to incentivize participants. Financial incentives have been shown to promote participants' adherence and enhance the effectiveness of interventions [96]. Studies targeting obesity or overweight populations for weight loss indicated that financial incentives can motivate healthy behaviors and facilitate weight loss [97-99].

Living Streaming

WeChat's live streaming function could encourage better engagement during health education. Because of the widespread transmission of the COVID-19 pandemic, people were required to quarantine at home and public places, such as parks and gyms, were closed, resulting in physical inactivity [100]. In view of this situation, the Chinese government expressed encouragement and support for live-streaming fitness in its National Fitness Plan (2021 - 2025) [101]. Genghong Liu's (a Taiwanese singer) free live-streamed exercise sessions on Douyin (the Chinese version of TikTok) attracted more than 13.9 billion viewers on April 17, 2022 [102]. This demonstrates the potential opportunity and serves as inspiration for using WeChat's live-streaming function to provide guidance on weight control.

Peer Support

Based on our review of intervention strategies in the included studies, WeChat group function served as a platform on which participants could communicate with each other to exchange experience and thoughts and provide peer support as well as encouragement [49,56,63,73,86]. Group support in a study using Noom, an app for obesity intervention, demonstrated positive impacts of peer support within the group, including enhanced health education acquisition (such as reading and responding to health-themed information) and increased responses to others' posts [103]. Evidence from Noom highlights the importance of social support, especially peer support, for future WeChat intervention studies.

Target Population

Although WeChat users are very diverse, older adults and young people, individuals with low education levels, and those living in poverty-stricken areas, many of whom are nonnetizens, still face difficulties using multiple WeChat functions or even accessing the platform [16]. Thus, implementing WeChat-based interventions among these populations might be limited.

Limitations

Our review has several limitations. First, methodological heterogeneity and variable reporting quality in Chinese language studies limited the direct comparability of outcomes, but we addressed this using narrative synthesis by study type (with or without a control group). Second, while the cultural specificity of WeChat-based interventions limits generalizability, this review lays important groundwork for cross-platform comparisons.

Conclusions

In summary, WeChat is a potential platform for delivering obesity interventions due to its wide use and various functions. Most of the reviewed papers reported effectiveness in terms of weight loss. However, intervention studies conducted via WeChat were limited in design and evaluation rigor. Therefore, future studies should include formative research for intervention development, use a randomized controlled blinded design, and

XSL•FO
RenderX

use objective obesity outcome measures to increase the reliability of their findings. The main features of WeChat used in the included obesity interventions enabled chatting and information sharing without spatial limitation. Thus, further studies should explore multiple functions simultaneously to enhance participant engagement with health education materials, examine the impact of varying exposure levels, and increase participant engagement.

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.

Data Availability

No new datasets were created in this scoping review. The manuscript synthesizes data from previously published studies (44-86).

Authors' Contributions

JG was in charge of the investigation process, project administration, and supervision. JG and YW conceptualized the study. YW and XZ worked on the data curation. YW conducted the formal analysis, while JG and YW worked on the design of methodology. YW worked on the visualization and original draft. All the authors contributed to the manuscript review and editing. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms and quality assessment.
[DOCX File, 45 KB - i-jmr_v14i1e65538_app1.docx]

Checklist 1
PRISMA-ScR checklist.
[DOCX File, 110 KB - i-jmr_v14i1e65538_app2.docx]

References

1. Congdon P. Obesity and urban environments. Int J Environ Res Public Health 2019 Feb 5;16(3):464. [doi: 10.3390/ijerph16030464] [Medline: 30764541]
2. Lee EY, Yoon KH. Epidemic obesity in children and adolescents: risk factors and prevention. Front Med 2018 Dec;12(6):658-666. [doi: 10.1007/s11684-018-0640-1] [Medline: 30280308]
3. NHC. Report on Chinese residents' chronic diseases and nutrition 2020. Ying Yang Xue Bao 2020;42(6):521 [FREE Full text]
4. Chen K, Shen Z, Gu W, et al. Prevalence of obesity and associated complications in China: a cross-sectional, real-world study in 15.8 million adults. Diabetes Obes Metab 2023 Nov;25(11):3390-3399. [doi: 10.1111/dom.15238] [Medline: 37589256]
5. Global health data exchange. : Institute for Health Metrics and Evaluation; 2019 URL: https://vizhub.healthdata.org/gbd-results?params=gbd-api-2019-permalink/dd64ad7d9c3fb70c835cbe706bc3fd9d [accessed 2024-04-08]
6. Wang Y, Zhao L, Gao L, Pan A, Xue H. Health policy and public health implications of obesity in China. Lancet Diabetes Endocrinol 2021 Jul;9(7):446-461. [doi: 10.1016/S2213-8587(21)00118-2]
7. Healthy china initiative (2019–2030). National Health Commission of the People's Republic of China. 2019. URL: https://www.nhc.gov.cn/guihuaxxs/c100133/201907/2a6ed52f1c264203b5351bdbbadd2da8.shtml [accessed 2025-10-05]
8. Roberto CA, Swinburn B, Hawkes C, et al. Patchy progress on obesity prevention: emerging examples, entrenched barriers, and new thinking. Lancet 2015 Jun 13;385(9985):2400-2409. [doi: 10.1016/S0140-6736(14)61744-X] [Medline: 25703111]
9. Hall KD, Kahan S. Maintenance of lost weight and long-term management of obesity. Med Clin North Am 2018 Jan;102(1):183-197. [doi: 10.1016/j.mcna.2017.08.012] [Medline: 29156185]
10. Perdomo CM, Cohen RV, Sumithran P, Clément K, Frühbeck G. Contemporary medical, device, and surgical therapies for obesity in adults. Lancet 2023 Apr 1;401(10382):1116-1130. [doi: 10.1016/S0140-6736(22)02403-5] [Medline: 36774932]
11. Dumas AA, Lapointe A, Desroches S. Users, uses, and effects of social media in dietetic practice: scoping review of the quantitative and qualitative evidence. J Med Internet Res 2018 Feb 20;20(2):e55. [doi: 10.2196/jmir.9230] [Medline: 29463487]

XSL•FO
RenderX

12. Bardus M, Smith JR, Samaha L, Abraham C. Mobile phone and Web 2.0 technologies for weight management: a systematic scoping review. J Med Internet Res 2015;17(11):e259. [doi: 10.2196/jmir.5129]
13. Tully L, Sorensen J, O'Malley G. Pediatric weight management through mHealth compared to face-to-face care: cost analysis of a randomized control trial. JMIR Mhealth Uhealth 2021 Sep 14;9(9):e31621. [doi: 10.2196/31621] [Medline: 34519665]
14. Woo Baidal JA, Chang J, Hulse E, Turetsky R, Parkinson K, Rausch JC. Zooming toward a telehealth solution for vulnerable children with obesity during Coronavirus disease 2019. Obesity (Silver Spring) 2020 Jul;28(7):1184-1186. [doi: 10.1002/oby.22860] [Medline: 32352650]
15. Robbins R, Krebs P, Jagannathan R, Jean-Louis G, Duncan DT. Health app use among US mobile phone users: analysis of trends by chronic disease status. JMIR Mhealth Uhealth 2017 Dec 19;5(12):e197. [doi: 10.2196/mhealth.7832] [Medline: 29258981]
16. The 53rd China Statistical Report on internet development. : CNNIC; 2024 Mar 22 URL: https://cnnic.cn/NMediaFile/2024/0325/MAIN1711355296414FIQ9XKZV63.pdf [accessed 2025-10-05]
17. Ji X, Hu L, Wang Y, et al. "Mobile Health" for the management of spondyloarthritis and its application in China. Curr Rheumatol Rep 2019 Nov 19;21(11):61. [doi: 10.1007/s11926-019-0860-7] [Medline: 31741083]
18. 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]
19. Siriwoen R, Chongsuwat R, Tansakul S, Siri S. Effectiveness of a weight management program applying mobile health technology as a supporting tool for overweight and obese working women. Asia Pac J Public Health 2018 Sep;30(6):572-581. [doi: 10.1177/1010539518800367]
20. Toro-Ramos T, Lee DH, Kim Y, et al. Effectiveness of a smartphone application for the management of metabolic syndrome components focusing on weight loss: a preliminary study. Metab Syndr Relat Disord 2017 Nov;15(9):465-473. [doi: 10.1089/met.2017.0062] [Medline: 29035677]
21. Wang E, Abrahamson K, Liu PJ, Ahmed A. Can mobile technology improve weight loss in overweight adults? A systematic review. West J Nurs Res 2020 Sep;42(9):747-759. [doi: 10.1177/0193945919888224] [Medline: 31762402]
22. Haas K, Hayoz S, Maurer-Wiesner S. Effectiveness and feasibility of a remote lifestyle intervention by dietitians for overweight and obese adults: pilot study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e12289. [doi: 10.2196/12289] [Medline: 30973338]
23. McDiarmid S, Harvie M, Johnson R, et al. Manchester Intermittent versus Daily Diet App Study (MIDDAS): a pilot randomized controlled trial in patients with type 2 diabetes. Diabetes Obes Metab 2022 Mar;24(3):432-441. [doi: 10.1111/dom.14592] [Medline: 34726317]
24. Huntriss R, Haines M, Jones L, Mulligan D. A service evaluation exploring the effectiveness of a locally commissioned tier 3 weight management programme offering face‐to‐face, telephone and digital dietetic support. Clin Obes 2021 Jun;11(3). [doi: 10.1111/cob.12444] [Medline: 33600056]
25. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz M, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019 Jan 8;21(1):e298. [doi: 10.2196/jmir.9609] [Medline: 30622090]
26. Cueto V, Wang CJ, Sanders LM. Impact of a mobile app-based health coaching and behavior change program on participant engagement and weight status of overweight and obese children: retrospective cohort study. JMIR Mhealth Uhealth 2019 Nov 15;7(11):e14458. [doi: 10.2196/14458] [Medline: 31730041]
27. 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]
28. Jelalian E, Darling K, Foster GD, Runyan T, Cardel MI. Effectiveness of a scalable mHealth intervention for children with overweight and obesity. Child Obes 2023 Dec;19(8):552-559. [doi: 10.1089/chi.2022.0154] [Medline: 36576892]
29. 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):e12990. [doi: 10.1111/ijpo.12990] [Medline: 36484235]
30. Wang Y, Willis E. Examining theory-based behavior-change constructs, social interaction, and sociability features of the weight watchers' online community. Health Educ Behav 2016 Dec;43(6):656-664. [doi: 10.1177/1090198116629415] [Medline: 27162239]
31. Thomas JG, Raynor HA, Bond DS, et al. Weight loss in Weight Watchers Online with and without an activity tracking device compared to control: a randomized trial. Obesity (Silver Spring) 2017 Jun;25(6):1014-1021. [doi: 10.1002/oby.21846] [Medline: 28437597]
32. Godino JG, Merchant G, Norman GJ, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755. [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
33. Rivera J, McPherson A, Hamilton J, et al. Mobile apps for weight management: a scoping review. JMIR Mhealth Uhealth 2016 Jul 26;4(3):e87. [doi: 10.2196/mhealth.5115] [Medline: 27460502]
34. Tencent announces 2024 first quarter results. Tencent. 2024 May 14. URL: https://www.tencent.net.cn/investors/results/ [accessed 2025-10-05]

XSL•FO
RenderX

35. Liu J, Zheng X, Zhang X, Feng Z, Song M, Lopez V. The evidence and future potential of WeChat in Providing support for Chinese parents of pediatric patients undergoing herniorrhaphy. J Transcult Nurs 2020 Mar;31(2):114-120. [doi: 10.1177/1043659619841672] [Medline: 30983517]
36. Li T, Ding W, Li X, Lin A. Mobile health technology (WeChat) for the hierarchical management of community hypertension: protocol for a cluster randomized controlled trial. Patient Prefer Adherence 2019;13:1339-1352. [doi: 10.2147/PPA.S215719] [Medline: 31496664]
37. Li WW, Toh P. WeChat-based intervention for Chinese immigrants with hypertension: development and evaluation study. Asian Pac Isl Nurs J 2023 Jul 27;7:e45769. [doi: 10.2196/45769] [Medline: 37498663]
38. Wang Y, Guo F, Wang J, et al. Efficacy of a WeChat-based multimodal digital transformation management model in new-onset mild to moderate hypertension: randomized clinical trial. J Med Internet Res 2023 Dec 4;25:e52464. [doi: 10.2196/52464] [Medline: 38048156]
39. Dong Y, Wang P, Dai Z, et al. Increased self-care activities and glycemic control rate in relation to health education via Wechat among diabetes patients: a randomized clinical trial. Medicine (Baltimore) 2018 Dec;97(50):e13632. [doi: 10.1097/MD.0000000000013632] [Medline: 30558051]
40. Ding B, Gou B, Guan H, Wang J, Bi Y, Hong Z. WeChat-assisted dietary and exercise intervention for prevention of gestational diabetes mellitus in overweight/obese pregnant women: a two-arm randomized clinical trial. Arch Gynecol Obstet 2021 Sep;304(3):609-618. [doi: 10.1007/s00404-021-05984-1] [Medline: 33570656]
41. Tian Y, Zhang S, Huang F, Ma L. Comparing the efficacies of telemedicine and standard prenatal care on blood glucose control in women with gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 25;9(5):e22881. [doi: 10.2196/22881] [Medline: 33783365]
42. Yang J, Yang H, Wang Z, et al. Self-management among type 2 diabetes patients via the WeChat application: a systematic review and meta-analysis. J Clin Pharm Ther 2021 Feb;46(1):4-16. [doi: 10.1111/jcpt.13264] [Medline: 33085789]
43. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89. [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
44. Study quality assessment tools. National Heart, Lung, and Blood Institute. 2013. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2025-10-05]
45. Xia SF, Maitiniyazi G, Chen Y, et al. Web-based TangPlan and WeChat combination to support self-management for patients with type 2 diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2022 Mar 30;10(3):e30571. [doi: 10.2196/30571] [Medline: 35353055]
46. He C, Wu S, Zhao Y, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017 Oct 23;19(10):e357. [doi: 10.2196/jmir.7861] [Medline: 29061555]
47. Chen JL, Guo J, Zhong Q, et al. Smartphone-Based Cancer and Obesity Prevention Education Program for Chinese Women (SCOPE): a pilot RCT. Int J Environ Res Public Health 2023 May 9;20(10):5768. [doi: 10.3390/ijerph20105768] [Medline: 37239495]
48. Liu Y, Sun P, Shuai P, Qiao Q, Li T. Fat-restricted low-glycemic index diet controls weight and improves blood lipid profile: a pilot study among overweight and obese adults in Southwest China. Medicine (Baltimore) 2021 May 28;100(21):e26107. [doi: 10.1097/MD.0000000000026107] [Medline: 34032752]
49. Li J, Yang H, Song X, et al. Effectiveness of social media with or without wearable devices to improve physical activity and reduce sedentary behavior: a randomized controlled trial of Chinese postgraduates. Heliyon 2023 Oct;9(10):e20400. [doi: 10.1016/j.heliyon.2023.e20400] [Medline: 37767499]
50. Han Y, Ye X, Li X, et al. Comparison of an online versus conventional multidisciplinary collaborative weight loss programme in type 2 diabetes mellitus: a randomized controlled trial. Int J Nurs Pract 2023 Feb;29(1):e13126. [doi: 10.1111/ijn.13126] [Medline: 36567135]
51. Ling D, Wang R, Chen Q, et al. Assessment of chronic disease management mode (CDMM) on participants with primary hypertension. Trop Med Int Health 2021 Jul;26(7):829-837. [doi: 10.1111/tmi.13577] [Medline: 33780099]
52. Yang X, Chattopadhyay K, Hubbard R, Li JL, Li L, Lin Y. 36-Month evaluation of a weight management programme in Chinese overweight and obese adults. Front Public Health 2021;9:749302. [doi: 10.3389/fpubh.2021.749302] [Medline: 34746084]
53. Linjuan Z. Analysis of risk factors of cardiovascular disease in type 2 diabetes patients and intervention of intelligent remote blood glucose management [Master's thesis]. : Inner Mongolia Medical University; 2023 URL: https://link.oversea.cnki.net/doi/10.27231/d.cnki.gnmyc.2023.000231 [accessed 2024-04-22]
54. Zihao P, Qi W, Mu C, et al. Analysis of weight loss effect of personalized high protein dietary intervention in overweight/obese people. Appl Prev Med 2022;28(4):312-315 [FREE Full text]
55. Xi C, Juan H, Hong L, et al. The application of 5+2 intermittent fasting in weight loss of employees in an affiliated hospital of a medical university. Med Diet Health 2021;19(3):25-27 [FREE Full text]
56. Dongmei X, Hui L, Danfeng C, et al. Application of Internet in health education for patients with coronary heart disease. J Qilu Nurs 2016;22(5):102-103 [FREE Full text]

XSL•FO
RenderX

57. Bei H. The application of mobile health treatment based on “internet +” for the health management of hypertension patients in a community [Master’s thesis]. : Xinjiang Medical University; 2018 URL: https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD201802&filename=1018212801.nh&uniplatform=OVERSEA&v=y_mpMf-c8AC-34R5kOgBNFkNgMHjc6We4usAY6waBSXKjhvIaEoNdgxGyD0Iemq2 [accessed 2024-04-22]
58. Xiaoyuan H, Meirong Z, Zhifang T, et al. Application of traditional Chinese medicine dietary therapy combined with WeChat exercise in nursing of coronary heart disease patients with phlegm dampness constitution. Dang Dai Hu Shi 2022;29(11):83-86 [FREE Full text]
59. Jinhua Y. Based on multi-theory model the study on comprehensive intervention of behavior changes for elderly patients with hypertension [phd thesis]. : Hunan University of Chinese Medicine; 2022 URL: https://link.oversea.cnki.net/doi/10.27138/d.cnki.ghuzc.2022.000074 [accessed 2024-04-22]
60. Jiawei C. Clinical study on the influence of traditional chinese medicine behavior intervention and health management on the quality of life of type 2 diabetes patients [Master’s thesis]. : Chengdu University of Traditional Chinese Medicine; 2020 URL: https://link.oversea.cnki.net/doi/10.26988/d.cnki.gcdzu.2020.000194 [accessed 2024-04-22]
61. Yuepeng T. A comparative study of online and offline exercise health education on behavioral changes in patients with hyperuricemia [Master’s thesis]. : Tianjin Institute of Physical Education; 2021 URL: https://link.oversea.cnki.net/doi/10.27364/d.cnki.gttyy.2021.000133 [accessed 2024-04-22]
62. Li Y. Design and application of internet plus nutrition and weight management system for diabetic patients [Master’s thesis]. : Army Medical University; 2018 URL: https://link.oversea.cnki.net/doi/10.27001/d.cnki.gtjyu.2020.000291 [accessed 2024-04-22]
63. Yingming M. Effect of health belief model on weight management in overweight or obese polycystic ovary syndrome patients with infertility [Master’s thesis]. : Hainan Medical University; 2020 URL: https://link.oversea.cnki.net/doi/10.27952/d.cnki.ghnyx.2020.000046 [accessed 2024-04-22]
64. Xiaojuan Y, Qun L, Mingzhu C, et al. Effect of individualized diet and exercises based on the WeChat on the overweight population with prediabetes. Nurs Integr Tradit Chin West Med 2018;04(11):20-24 [FREE Full text]
65. Peijun H. Effect of traditional Chinese medicine health care guidance on the WeChat platform on body weight and quality of life for patients with obesity. J Sichuan Tradit Chin Med 2022;40(2):219-222 [FREE Full text]
66. Yuexi Y, Yanxia L, Xiao W, et al. Effect of WeChat multimedia platform on the glucose metabolism in pre diabetes population. J SNAKE Sci Nat 2019;31(4):473-476 [FREE Full text]
67. Jiangping M, Qiuli Z, Shanshan C, et al. The effectiveness of exercise intervention on exercise adherence among college teachers with dyslipidemia. Chin J Nurs Educ 2016;13(6):420-425 [FREE Full text]
68. Xiao Z, Dafang C. Effects of lifestyle intervention by internet-based online on weight loss in people with weight loss willingness. Chin J Prev Contr Chron Dis 2019;27(4):257-260 [FREE Full text]
69. Tianmeng R, Jie R, Zilong L, et al. Effects of short-term walking intervention on physical health indicators in occupational population with higher blood pressure. Chin J Prev Contr Chron Dis 2023;31(4):256-261 [FREE Full text]
70. Ziwei Y. Effects of wechat intervention based on BCW+bcts on the treatment compliance of young and middle-aged patients with hypertension [Master’s thesis]. : University of South China; 2021 URL: https://link.oversea.cnki.net/doi/10.27234/d.cnki.gnhuu.2021.000792 [accessed 2024-04-22]
71. Jianxin M, Jinping Z, Lian C, et al. Efficiency of health management for elderly hypertensive patients using WeChat Applet. Chin J Mult Organ Dis Elderly 2020;19(12):895-899 [FREE Full text]
72. Di P. Energy balance and imbalance of obese children’s weight changes—an intervention program on diet and exercise [phd thesis]. : Beijing Sport University; 2017 URL: https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2017&filename=1017181664.nh&uniplatform=OVERSEA&v=KgfQHfMuKQK2t_YxPsLSMI5ih1thlN_qdktKt8v4-MYCTcBCi7dVsvMwGkS0A-Ep [accessed 2024-04-22]
73. Ruixue B, Xiaoya Q, Yukun H, et al. Evaluation of mobile application-based physician active management on weight loss in overweight and obese patients. Health Med Res Prac 2023;20(1):01-05 [FREE Full text]
74. Youxuan Y. Evaluation of the effect of internet plus intervention on blood pressure control in high risk population of hypertension in haikou community [master’s thesis]. : Hainan Medical College; 2023 URL: https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202401&filename=1023643307.nh&uniplatform=OVERSEA&v=aTZXVkrghMBTfdvmOY4tHZLSk_u6Qm2KuTOkmTBO3dSsMo6pPUO-u2ulv3szPLG9 [accessed 2024-04-22]
75. Jielili M. Evaluation of the weight loss effect of whole wheat flour on obese people [Master’s thesis]. : Xinjiang Medical University; 2022 URL: https://link.oversea.cnki.net/doi/10.27433/d.cnki.gxyku.2020.000396 [accessed 2024-04-22]
76. Mengqin W. Health management and its evaluation of essential hypertension among residents in underdeveloped rural areas of shanxi province [Master’ thesis]. : Shanxi Medical University; 2023 URL: https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202401&filename=1023825479.nh&uniplatform=OVERSEA&v=EYMG9C7Wn20OxkBNdrwiNNz_kZKLad7v2IaZeI-njwMio3ChEKPxbbFEWt7or1wQ [accessed 2024-04-22]

XSL•FO
RenderX

77. Xujuan M, Junfeng Z, Minghou Z, et al. The impact of 8-week health management on fat mass and other indicators in obesity population. J Shanxi Med Coll Contin Educ 2019;29(3):93-96 [FREE Full text]
78. Wei Y, Yongyan Y, Xiaokang H, et al. The impact of health education on WeChat platform on the health knowledge and behavior of college students with normal high blood pressure. Xizang Med 2023;44(1):150-152 [FREE Full text]
79. Rongrong L. Management of patients with prediabetes based on wechat platform and its effect analysis [Master's thesis]. : Xinjiang Medical University; 2019 URL: https://link.oversea.cnki.net/doi/10.27433/d.cnki.gxyku.2019.000060 [accessed 2024-04-22]
80. Huirong G, Xueqin W, Wei L, et al. Research of intervention by "public health microworld" WeChat platform for lifestyle. China Prac Med 2017;12(15):184-187 [FREE Full text]
81. Weiwei C. Research on the intervention effect of aerobic exercise program on the physical inactivity of the elderly patients with chronic diseases based on COM-b model [Master's thesis]. : Bengbu Medical College; 2022 URL: https://link.oversea.cnki.net/doi/10.26925/d.cnki.gbbyc.2022.000289 [accessed 2024-04-22]
82. Moufu W, Xuejiao Z, Ting L, et al. Research on the management effect of "Internet +" self-management mode on diabetes mellitus patients. China Prac Med 2022;17(24):180-183 [FREE Full text]
83. Lu Q. Study on the application of calorie restrict diet intervention based on IMB model in young and middle-aged patients with coronary heart disease [master's thesis]. : North China University of Science and Technology; 2023 URL: https://link.oversea.cnki.net/doi/10.27108/d.cnki.ghelu.2023.000895 [accessed 2024-04-22]
84. Yinbao S. Study on the effect of health education on cardiovascular diseases among Uyghur population in rural areas of Xinjiang production and construction corps [Master's thesis]. : Shihezi University; 2023 URL: https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202401&filename=1023708617.nh&uniplatform=OVERSEA&v=DfrWkfZQKBjDQeEDJu77c2GIA4DmAsqAcAOKzlc7kL0GTwVfntpWdPfkwrI-E7DB [accessed 2024-04-22]
85. Lijuan Z. Study on the evaluation of intervention effects of the O2O-based community management model for young and middle-aged hypertension patients [master's thesis]. : Fujian Medical University; 2020 URL: https://link.oversea.cnki.net/doi/10.27020/d.cnki.gfjyu.2020.000195 [accessed 2024-04-22]
86. Jun L. Study on the intervention effect of eight- form tai chi on pre-diabetic people under different exercise management mode [Master's thesis]. : Guangzhou Sport University; 2023 URL: https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CMFD&dbname=CMFD202401&filename=1023765390.nh&uniplatform=OVERSEA&v=W1Vtw8Qsgx4EWVhrd4mflQb1RkwPYXr7mWjbCKHCKylvMcYgJhL9MJGJ_Ui9opnP [accessed 2024-04-22]
87. Jingxia K, et al. Worksite-based intervention on prevention and control of overweight and obesity [phd thesis]. : Zhejiang University; 2020 URL: https://oversea.cnki.net/KCMS/detail/detail.aspx?dbcode=CDFD&dbname=CDFDLAST2021&filename=1020435505.nh&uniplatform=OVERSEA&v=CexL3PgZ7pMiYuDn4f68lVAyrrJMoMdLTzzVAfAEE1UX0mb2H9BVIzVX11AfBpMt [accessed 2024-04-22]
88. Gittelsohn J, Steckler A, Johnson CC, et al. Formative research in school and community-based health programs and studies: "state of the art" and the TAAG approach. Health Educ Behav 2006 Feb;33(1):25-39. [doi: 10.1177/1090198105282412] [Medline: 16397157]
89. Bentley ME, Johnson SL, Wasser H, et al. Formative research methods for designing culturally appropriate, integrated child nutrition and development interventions: an overview. Ann N Y Acad Sci 2014 Jan;1308:54-67. [doi: 10.1111/nyas.12290] [Medline: 24673167]
90. Gupta L, Lal PR, Gupta Y, Goyal A, Khanna A, Tandon N. Formative research to develop diabetes self-management education and support (DSMES) program for adults with type 1 diabetes. Diabetes Metab Syndr 2021;15(4):102150. [doi: 10.1016/j.dsx.2021.05.023] [Medline: 34186364]
91. Adebile TV, Adebile TM, Oloyede TF, et al. Telemedicine for obesity management among United States adults: a systematic and meta-analysis of intervention studies. J Telemed Telecare 2025 Sep;31(8):1078-1090. [doi: 10.1177/1357633X241247240] [Medline: 38632958]
92. James A, Afable A, Bayoumi N, Dhuper S. Evaluation of a childhood obesity program serving a high-need population in Brooklyn, New York using survival analysis. Int J Environ Res Public Health 2023 May 4;20(9):5723. [doi: 10.3390/ijerph20095723] [Medline: 37174241]
93. Yun K, Chu Z, Zhang J, et al. Mobile phone intervention based on an HIV risk prediction tool for HIV prevention among men who have sex with men in China: randomized controlled trial. JMIR Mhealth Uhealth 2021 Apr 13;9(4):e19511. [doi: 10.2196/19511] [Medline: 33847597]
94. Shangguan F, Wang R, Quan X, et al. Association of stress-related factors with anxiety among Chinese pregnant participants in an online crisis intervention during COVID-19 epidemic. Front Psychol 2021;12:633765. [doi: 10.3389/fpsyg.2021.633765] [Medline: 33995188]
95. Song Y, Xie X, Chen Y, et al. The effects of WeChat-based educational intervention in patients with ankylosing spondylitis: a randomized controlled trail. Arthritis Res Ther 2021 Mar 4;23(1):72. [doi: 10.1186/s13075-021-02453-7] [Medline: 33663558]

XSL•FO
RenderX

96. Voils CI, Pendergast J, Hale SL, et al. A randomized feasibility pilot trial of a financial incentives intervention for dietary self-monitoring and weight loss in adults with obesity. Transl Behav Med 2021 Apr 26;11(4):954-969. [doi: 10.1093/tbm/ibaa102] [Medline: 33245118]
97. Agarwal AK, Waddell KJ, Small DS, et al. Effect of gamification with and without financial incentives to increase physical activity among veterans classified as having obesity or overweight: a randomized clinical trial. JAMA Netw Open 2021 Jul 1;4(7):e2116256. [doi: 10.1001/jamanetworkopen.2021.16256] [Medline: 34241628]
98. Hoddinott P, O'Dolan C, Macaulay L, et al. Text messages with financial incentives for men with obesity: a randomized clinical trial. JAMA 2024 Jul 2;332(1):31-40. [doi: 10.1001/jama.2024.7064] [Medline: 38744430]
99. Yeung KF, Gandhi M, Lam AYR, et al. The Pre-Diabetes Interventions and Continued Tracking to Ease-out Diabetes (Pre-DICTED) program: study protocol for a randomized controlled trial. Trials 2021 Aug 6;22(1):522. [doi: 10.1186/s13063-021-05500-5] [Medline: 34362409]
100. Tian R, Yin R, Gan F. Exploring public attitudes toward live-streaming fitness in China: a sentiment and content analysis of China's social media Weibo. Front Public Health 2022;10:1027694. [doi: 10.3389/fpubh.2022.1027694] [Medline: 36408048]
101. The National Fitness Plan (2021-2025). The State Council of the People's Republic of China. 2021. URL: https://www.gov.cn/zhengce/content/2021-08/03/content_5629218.htm [accessed 2025-10-05]
102. Liu Genghong, dubbed "next Li Jiaqi", unlocks China's fitness live stream. Dao Insights. 2022. URL: https://daoinsights.com/news/liu-genghong-dubbed-next-li-jiaqi-unlocks-chinas-fitness-live-stream/ [accessed 2025-10-05]
103. Kim H, Ray CD, Veluscek AM. Complementary support from facilitators and peers for promoting mHealth engagement and weight loss. J Health Commun 2017 Nov 2;22(11):905-912. [doi: 10.1080/10810730.2017.1373876]

Abbreviations

BF: body fat
BFP: body fat percentage
CNKI: China National Knowledge Infrastructure
HC: hip circumference
mHealth: mobile health
NCD: noncommunicable chronic diseases
NIH: National Institutes of Health
NR: not report
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
WC: waist circumference
WHR: waist-to-hip ratio

Edited by TDA Cardoso; submitted 21.08.24; peer-reviewed by J Chan, K Yun; revised version received 30.06.25; accepted 11.07.25; published 11.12.25.

Please cite as:
Wang Y, Zhuang X, Sundermeir S, Gittelsohn J
Obesity Prevention and Reduction in China Using the Social Media Platform WeChat: Scoping Review
Interact J Med Res 2025;14:e65538
URL: https://www.i-jmr.org/2025/1/e65538
doi:10.2196/65538

© Yinuo Wang, Xuxiu Zhuang, Samantha Sundermeir, Joel Gittelsohn. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 11.12.2025. 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

Developing Requirements for a Standardized System to Return Individual Research Results Back to Study Participants: Narrative Review

Rosalyn Leigh Carr[1,2], BASc; Vita Chan[2,3], MLIS; Nicholas C West[2], MSc; Matthias Görges[2,4], PhD

[1]The School of Biomedical Engineering, University of British Columbia, 6088 University Boulevard, Vancouver, BC, Canada

[2]Research Institute, BC Children's Hospital, Vancouver, BC, Canada

[3]School of Information, University of British Columbia, Vancouver, BC, Canada

[4]Department of Anesthesiology, Pharmacology and Therapeutics, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Rosalyn Leigh Carr, BASc
The School of Biomedical Engineering, University of British Columbia, 6088 University Boulevard, Vancouver, BC, Canada

Abstract

Background: The increasing prevalence of smart devices has created vast amounts of untapped data, presenting new opportunities for data sharing across various fields, such as environmental sciences, health management, and astrophysics. While a significant portion of the public is willing to donate personal data, we need to better understand how to obtain information about which data assets a person may hold and the risks, benefits, and potential uses of this data exchange mechanism. Developing a trusted data-sharing platform may increase participants' willingness to donate data and researchers' ability to return personalized results from research findings.

Objective: This study aimed to develop a preliminary list of core requirements, which can be used to develop design recommendations for standardizing the return of individual research results to study participants across research disciplines.

Methods: We conducted a narrative literature review of existing platforms used to return research results to study participants. The search strategy included English-language articles published between May 2013 and May 2023. Concepts related to returning, disseminating, and sharing research results were searched for in (1) published research reports on Web of Science and MEDLINE, (2) gray literature, and (3) the bibliographies of included articles. Screening and data extraction were performed by 2 independent reviewers using Covidence. Inclusion criteria required that the study (1) included human participants, (2) returned information based on data collected from or by participants, (3) was published in English, and (4) included a description of a results-sharing system. Articles that met all 4 inclusion criteria were included in the review; articles that met the first 3 were also presented as supplementary articles. Results and requirements were synthesized thematically.

Results: Overall, 6608 abstracts were screened, and 266 articles underwent full-text review to identify 8 articles describing the development and evaluation of 7 different return of results systems. In total, 7 of the 8 articles reported the use of multimodal dissemination methods, including a combination of physical documents, emails, phone calls, and digital platforms to support text and graphical data representations. One article outlined accessibility features to serve the specific participant population. None of the articles described in detail how results were or were not anonymized. A total of 4 studies relied on an expert or clinician to share results on behalf of the research team. Additional educational or contextual materials were included alongside results in four studies, including specific materials designed for follow-up with experts and clinicians. Participants were not hesitant to receive unfavorable results and instead aimed to incorporate such information into their lives via lifestyle changes, clinical intervention, or seeking community.

Conclusions: Return of results systems should support multiple modes of dissemination for text-based results. Additional educational and lay-language materials are helpful for participants to understand and use information gained from receiving results.

(Interact J Med Res 2025;14:e65606) doi:10.2196/65606

KEYWORDS

return of results; research results; knowledge translation; system; researching; researcher; research participants; study participants; personalization; data integration; data visualization; research ethics; narrative review

XSL•FO
RenderX

Introduction

Background

In 2019, it was estimated that almost 70% of households had a smart device or appliance [1], and as of 2021, 85% of North Americans owned and used a smartphone [2]. This proliferation of smart technology has generated vast amounts of largely untapped data, presenting tremendous new opportunities for open science and knowledge sharing—particularly in areas where citizen scientists [3] want to contribute to academic discovery and applied research. Examples include water quality [4], astrophysics [5], environmental sciences [6], and patient-generated health and activity data [7], which can affect patients' health management by enhancing their ability to convey their health status [8] and improving patient-clinician relationships [9].

As with biobanks, where people can donate samples of tissues or body fluids for targeted or unrestricted use in research and health care [10], there is growing interest in donating personal data in various research areas [11]. Examples include a Swedish activity and physiology data sharing app [12], the ecobee smart thermostat "donate your data" dataset [13], or Tidepool's continuous glucose monitor "Big Data Donation Project" [14]. Increasing public trust may expand the use of this approach.

There is growing support for citizens to be able to donate their data for research. In a United Kingdom survey of 1300 members of the public, 54% of respondents were willing to donate personal data for research [15], motivated by both self-benefit and altruism [15]. However, barriers to data donation include a need to better understand the risks and benefits associated with their data and its potential uses, as well as gain insights into which data assets a person may hold [16,17]. Combined with privacy concerns regarding data-sharing methods, this highlights the need to create trusted data-sharing platforms. Similarly, there may be an enhanced willingness to share data if there is perceived self-benefit, such as the ability to derive insights from one's data or benchmark results against others, which could help their interpretation and provide locally relevant context [18].

Acknowledging these challenges, the National Academies of Sciences, Engineering, and Medicine assembled a committee to develop comprehensive guidance on returning individualized research results from biospecimens [19]. This report outlines 12 key recommendations to support study-specific decision-making, enhance research quality, improve participant understanding, and align with current regulations.

The recommendations emphasize the significance of considering the value of results to participants, evaluating the quality of research laboratories, and engaging community and participant representatives in policy development [19]. They also advocate providing adequate resources for high-quality research, planning for result return in funding applications, and supporting research to broaden the empirical evidence base. In addition, the recommendations call for revising regulations to facilitate the return of individual research results, harmonizing definitions across federal regulations, and ensuring transparency during the consent process. Finally, while the report addresses various factors influencing the feasibility of returning individual results to participants, including the potential burdens that investigators face when doing this for a single study, it does not discuss the systems available for doing so at scale (ie, by establishing infrastructure to be leveraged across multiple studies) or their potential impact on the widespread adoption of practices to return individualized results to study participants.

Existing studies on the return of results have likewise focused on policy, guidance, and best practices [19-22], or the perceptions of both participants and researchers regarding the return of results [23-26]. However, there has not been a formal review of the tools and methods used to return results, as many studies are restricted solely to their relevant discipline and data. This study aims to expand on this discussion.

Some fields have existing context-specific platforms to return research results. Environmental sciences have the digital exposure report-back interface, which returns chemical exposure results, such as lead contamination in drinking water [27], and Gardenroots, which illustrates contaminant levels in water, soil, and plant samples contributed by participants [28]. There are recommendations for returning environmental health-related data results [29], but other fields have less well-defined mechanisms [30].

Objective

This study aimed to develop a preliminary list of core requirements, which can be used to develop design recommendations for the standardizing of the return of individual research results to study participants across research disciplines.

Methods

Study Design

We conducted a narrative review of the published literature for reports of existing platforms used to return research results to study participants. The protocol was registered at the Center for Open Sciences' registry on May 24, 2023 [31]. A narrative review was chosen because the breadth and heterogeneity of this topic made a systematic review impractical. Systematic reviews are ideal for narrowly focused questions and structured data synthesis. In contrast, narrative reviews allow for broader interpretation and integration of diverse sources, which was essential for our research aims [32]. We followed the registered protocol for article selection but extended it to present a larger set of articles that met a limited set of our inclusion criteria as supplementary material.

Literature Search Strategy

Terms related to research results, result dissemination, and sharing of individual data from studies were used to search the literature. The search method combined search terms using proximity searching to identify key phrases related to the methods and process of returning research results. This approach was chosen instead of traditional keyword searching as the terms describing the method of returning research results to participants were often not specific enough to avoid large

XSL•FO
RenderX

quantities of extraneous articles. These included cases of "research" being used as an adjective alongside "data" or "results," while similarly, the term "results" tended to include headers and titles within abstracts rather than methods. An initial search strategy was created with the guidance of a research librarian to identify key phrases to search in abstracts, titles, and keywords (Multimedia Appendix 1).

A search was performed on Web of Science (Appendix A) and MEDLINE (Ovid; Appendix B). All searches were limited to articles published between May 2013 and May 2023. Gray literature was also searched by including the first 100 results with the search term "returning research results to participants" on Google Scholar. IEEE Xplore was also searched using this phrase and others, as multiple proximity operators are not supported within the same search. Finally, the reference lists of all identified reports and articles were iteratively searched for additional studies until no more relevant articles were identified.

All articles were uploaded into Covidence (Veritas Health Innovation, Melbourne, Australia), and duplicates were removed.

Article Selection Process

Two reviewers (RC and VC) independently completed an initial screening based on the article titles and abstracts to identify which articles would be included in the full-text review. The inclusion criteria were: (1) the reported study included human participants (members of the public or nonexperts) receiving information on their study results; (2) the information returned to participants was based on data previously collected from or by participants and included at least one variable with a description of the variable (ie, what was being measured) and domain context (ie, the meaning of this measurement in the context in which the study was conducted); (3) the article was published in English; (4) the article must include a description of a system or design (implemented as a pilot, at least) for sharing research results. Articles that met all 4 inclusion criteria were included in the full review, while articles that met the first 3 inclusion criteria but not the fourth were included as supplementary articles.

Two reviewers (RC and VC) then independently performed a full-text screening of the selected articles. Disagreements were resolved through discussion; if there was no resolution, additional reviewers were added to achieve consensus. Articles meeting criteria (1), (2), and (3), but not (4), formed a larger pool of supplementary articles of secondary interest. Articles that met all 4 inclusion criteria, including the additional system criteria, were selected to undergo full data extraction.

Data Extraction

Two independent reviewers (RC and VC) used a standardized data extraction form in Covidence to extract data from each selected article (Multimedia Appendix 2). The following information was extracted into the template: study details, study characteristics, population characteristics, data sharing methodology, and data characteristics.

Risk of Bias and Quality Assessment

Two researchers (RC and VC) independently rated the quality of each study using the Authority, Accuracy, Coverage, Objectivity, Date, Significance checklist [33], as none of the included articles included randomized control trials, which was a prerequisite for using other tools.

Data Analysis

Data collected were analyzed and synthesized by 2 researchers (RC and VC) using thematic analysis [34]. Open coding was initially used to identify key segments in the data, after which themes were developed inductively through an iterative process. These themes were then organized into a hierarchy to form key findings. The findings include summarized information on study characteristics, variables measured, data presentation, delivery mechanism, results contextualization, whether the data shared are personalized, and system design requirements.

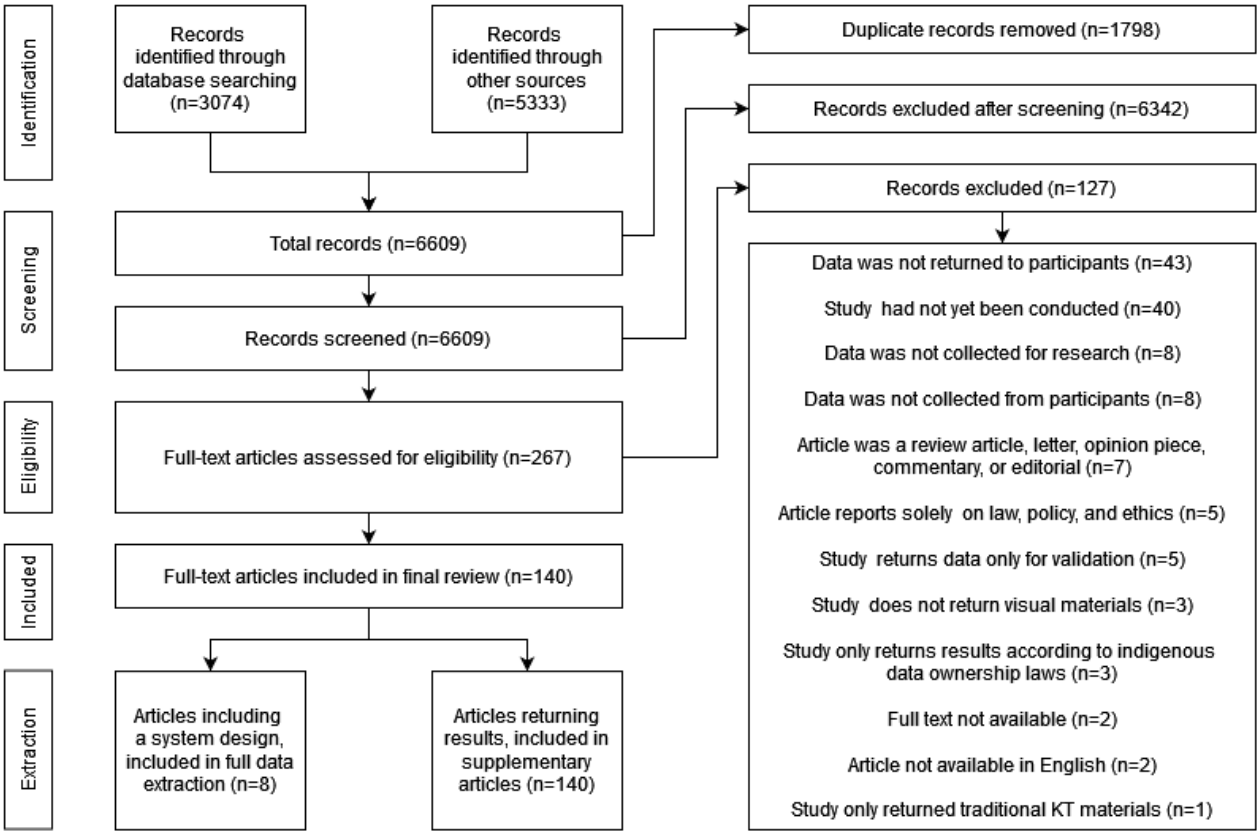
Results

Study Selection

After removing duplicates, 6609 articles were identified from the literature search and bibliographies of iteratively included papers. After reviewing the titles and abstracts, 267 articles were selected for full-text review. The total number of articles included in the supplementary list was 140 (Multimedia Appendix 3). From this, 8 articles [27,35-41] met all 4 inclusion criteria; however, 2 articles described a single study, discussing its design criteria in 1 article [27] and its implementation in another article [40], which were combined, resulting in a total of 7 studies (Figure 1).

XSL•FO
RenderX

Figure 1. Study selection and inclusion flow chart using Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) categories. All articles included in the full data extraction are also included in the supplementary articles, as they meet the inclusion criteria for both groups.



Study Characteristics

The eight included articles (Table 1) were published between 2017 and 2023, with 5 articles from the health care domain [35-39] and 3 (describing 2 studies) from the environmental health domain [27,40,41]. In total, 4 studies were based in the US [27,35,38,40,41], 1 in the Netherlands [39], 1 in the UK [37], and 1 in Japan [36]. The total number of study participants varied from 20 up to 24,163, with 1 study including 7 researchers as additional participants to provide feedback on system usability [41].

In total, 5 of the 7 studies involved vulnerable groups, including children [27,40,41], infants and mothers [38], retinal disease patients with significant vision loss [37], and cancer survivors [39]. In 2 studies, participants were identified and recruited from previously collected data (previous genetic testing or whole genome sequencing) [35,36]. In total, 4 studies included recruitment directly through a health care provider (HCP) or hospital setting [35,37-39], with 1 of these studies being a company recruiting via local HCPs [37]. In 1 study, the primary participants were researchers [41], and in 1 study, participants were recruited through community engagement or door-to-door soliciting [27,40]. Financial compensation was provided for participation in some study activities in 1 study [38].

XSL•FO
RenderX

Table . Study characteristics of the seven studies including system design.

Study	Study domain	Participant characteristics	Sample size	Country or region
Boronow 2017 [27]; Perovich, 2018 [40]	Environmental health	Children with doctor-diagnosed asthma	10 participants completed feedback interview	US; Metro-Boston (Lowell, Lawrence, Old Colony, Castle Square) and Cincinnati
Cope, 2023 [38]	Health care	Babies of birthing parents at least 11 years of age	1823 participants completed feedback survey, 24 participants completed feedback interview	North Carolina
Gilbert, 2022 [37]	Health care	Patients with inherited retina disease	20 participants attended initial focus groups, 30 participants attended technology workshops, 50 participants attended user research focus groups	UK; London
Ohneda, 2022 [36]	Health care	Population-based genome cohort	161 participants completed questionnaire survey before receiving results, 150 participants completed questionnaire survey after receiving results	Japan
Polka, 2021 [41]	Environmental health	Homes in two environmental justice communities, children in areas with manganese in their local drinking water	7 researchers interviewed	US; Massachusetts
van de Poll-Franse, 2022 [39]	Health care	Cancer patients or survivors	N/A[a] (population registry)	Netherlands
Savatt, 2018 [35]	Health care	Anyone who has had genetic testing regardless of genetic test results or diagnosis	1601 participants enrolled in GenomeConnect	US

[a]not applicable.

Risk of Bias and Quality Assessment

Based on the Authority, Accuracy, Coverage, Objectivity, Date, Significance checklist (Multimedia Appendix 4), none of the seven studies had a strong indication of bias, with 3 having minor concerns.

Method of Dissemination

Several modalities of returning results were reported, with seven articles describing the use of 2 or more methods. Results were returned as a physical report through traditional mail in 2 studies [36,40], and 1 also returned results at an in-person community meeting [40]. A total of 4 studies used an online web-based portal [35,36,38,39], and 1 used a dedicated mobile application with specific accessibility features for people with vision loss [37]. Finally, 1 study used an online portal that required a clinician to deliver specific results (such as positive screening results) by phone [36] or an unspecified mechanism [38]. In total, 2 studies required clinicians to review before dissemination, though the clinicians did not disseminate results directly [35,37]. Emails were used by 2 studies to notify participants that their results were available [35,38] or that changes had been made (such as a new interpretation of previously collected data) [35].

Data Format

Out of 7 studies, 6 studies returned at least part of the participants' results in a text format (written word) [27,35-38,40,41] (Table 2), with 1 article not specifying the data format used [39]. Visuals were also included in 4 studies as charts or graphs [27,40,41], image data (such as retinal scans) [37], or graphics (non-data images used for aesthetic purposes) [36]. A total of 3 studies created their own results report for their participants [27,36,40,41], 1 returned standard medical laboratory reports [38], and 1 created a specific "physician report" alongside the participant report for sharing with their HCP [36].

In addition, 1 study also included nontraditional results in the form of personalized t-shirts for participants that included the number of chemical values measured in their homes and aggregate results in the form of data physicalization [42], which participants could observe at a community meeting organized by researchers [40].

XSL•FO
RenderX

Table . Data characteristics of the seven studies including system design.

Study	Presentation of information	Mechanism of delivery	Contextualization of results	Identifiable	Education of results
Boronow 2018 [27]; Perovich, 2017 [40]	Two report-back packets designed with the DERBI system containing personalized graphs and text summaries, personalized data shirt, and data physicalization (physical objects that display data in the real world instead of on paper or a screen)	The two report-back packets were delivered in person during a home visit and a community meeting. Personalized data shirts and data physicalization were available to participants who could attend the community meetings. Participants who could not attend were mailed their report-back packets.	In addition to personalized results, participants received aggregate results of chemical levels in the homes of other participants in the study cohort.	Personalized	Yes
Cope, 2023 [38]	Portal results page explains the meaning of a normal result, accuracy of the screening test, and information about the diseases that were screened. CLIA-certified laboratory report provided.	Participants with normal results were notified via email when results were available on the portal. Participants with positive screening results were notified by the study genetic counselor.	No context	Personalized	Yes
Gilbert, 2022 [37]	Not reported	MyEyeSite patient application	Not reported	Personalized	
Ohneda, 2022 [36]	Participants received information about generic and brand names of drugs considered as potential risks, along with their indication, medicinal effects, and individual predicted genotyped-based response to the drug. Health professionals of patients were given a single sheet containing all three genes, medication, and dosage recommendations based on the genomic results	Mail and genetic counseling through telephone	No context	Personalized	Yes
Polka, 2021 [41]	Numerical, text, and graphical representations of participant's results	Not reported	Population and study cohort	Personalized	Yes
van de Poll-Franse, 2022 [39]	Not reported	Not reported	Population; age- and sex-matched normative population with cancer	Personalized	Yes
Savatt, 2018 [35]	Notification with general statement that genetic test results have been updated and suggestion to contact the ordering healthcare provider or genetics provider in their area	Email	Match with other participants with similar characteristics (eg, gene, disease, U.S. state)	Not reported	Not reported

XSL•FO
RenderX

Contextualization, Education, Recommendations, and Results Literacy

In 4 studies, additional details were provided alongside returned results to contextualize the information and provide recommendations [27,36,38,40,41]. One article (which described 2 pilot studies) described the return of individual in-home air concentrations of common contaminants, accompanied by general information on common pollutant sources and population averages for comparison [41]; the second pilot study returned individual tap water manganese levels, including general information and "group-level information," likely aggregated data from the study population, but which was not specified [41]. Another study included additional information on diseases for which participants had been screened in the results [38]. Other studies included recommendations about "improving" one's results (lowering exposure to specific chemicals) [40] or recommendations for avoiding specific drugs based on one's pharmacogenetic results within the results [36].

Expert Intervention, Follow-Up, and Community Engagement

Experts, primarily clinicians such as genetic counselors, were involved in 4 studies at various stages [35-38]. In 2 studies, clinicians reviewed results before dissemination [35,37], and 2 required a clinician to deliver specific results [36,38]. Genetic counseling was mandated for participants with specific results in one study [36]. Finally, 1 study, which allowed participants to upload data directly to the application, would automatically notify an HCP if data exceeded predefined thresholds so that an intervention could occur [37].

Follow-up with experts outside of the study team was also common: 2 studies nudged their participants to follow up with their primary HCP [35,36], one of which had prepared and included tailored results for this purpose [36]. Participants also expressed their plans to share their results with their HCPs [37,38,40]. However, this motivation was sometimes mediated by the results, as participants in 1 study did not feel the need to report "normal" results to experts [38].

When participants were not directly engaging with experts, one platform also provided methods to engage with additional research opportunities and clinical trials [37]. In contrast, another study looked at creating or maintaining a community involved in the research through community events and engagement sessions [40]. One platform also allowed participants to view relevant products to purchase [37].

Identifiability and Anonymity

The extent to which data or results were identifiable was not clearly stated in all 8 articles. Researchers in 1 study used participant IDs in the internal database that were only accessible by researchers to avoid linking identifiers [36]. Another system only made information accessible to researchers that could not be easily identifiable (such as unlabeled images or appointment schedules) [37]. The methods were unspecified in one article, though data were described as "anonymous" [35] and 2 articles did not specify any level of inherent anonymization unless researchers deliberately did this using outside tools or methods [27,41]. Finally, 1 article did not mention privacy or anonymity, but the registry was tied to the local government and health authorities [39].

Participant Perceptions

Participant perceptions of results were positive, with sentiments of increased trust, autonomy, and knowledge being key factors [27,37,39,40] (Table 3). In 1 study, participants did not experience worry or anxiety associated with receiving results [27,40]. Negative perceptions included a lack of adequate education materials or appropriate language [36,38], inadequate anonymization and options to opt-out of communications [37], or not having the option to receive additional results if they became available [35].

XSL•FO
RenderX

Table . Feedback and conclusions of the seven studies including system design.

Study	Positive participant opinions	Negative participant opinions	Researcher opinions
Boronow 2018 [27]; Perovich, 2017 [40]	Motivated to join because participants are able to access information that wouldn't be available otherwise. Participants felt like they were able to build trust with researchers. Results did not cause excess worry/anxiety/distress to participants. Participants made specific positive changes and brought study results to medical resources.	Participants recommend creating report-back materials for different audiences.	Digital exposure report-back interface is good for large studies where personalization is very important.
Cope, 2023 [38]	Calling participants with positive screening results ensures that participants with significant implications for their health are aware. The portal was rated easy to use and helpful for understanding individual research results. But literature indicates that online portals may be only acceptable to some participants.	Avoid technical terms and use plain language as much as possible.	Returning individual research results online was a time-efficient method that reduced burden on study staff.
Gilbert, 2022 [37]	Participants reported increased agency and being empowered to manage their own conditions. Information should be expressed in lay terms.	Patients who lack experience with digital tools rely on others for assistance. Data needs to be anonymized, and there should be easy opt-in/opt-out options for research studies and receiving notifications for eligible studies and findings from studies	—[a]
Ohneda, 2022 [36]	Participants' knowledge of PGx testing improved after receiving results	Even though the basic information of PGx had been given several times, some participants felt that the PGx concepts and terminology were difficult to fully understand	—
Polka, 2021 [41]	—	—	Manual report-back allows researchers to become familiar with participants' results, but the process is time-intensive, which limits individualization and creativity and presents challenges of data accuracy
van de Poll-Franse, 2022 [39]	Return of individual results not only empowers patients but can also be an incentive for cohort retention. Comparison to an age- and sex-matched normative population is helpful in the interpretation of findings	—	—
Savatt, 2018 [35]	—	Registry participants want to receive information about potential updates to their genetic testing results (ie, if the interpretation has changed)	—

[a]not applicable.

Researcher Perceptions

There was 1 study that included interviews with the researchers who were tasked with creating and disseminating the results. Even with a dedicated system, researchers report that the process was time-consuming, particularly for individual results [41]. Another study that included interviews with the research team as well as the participants who received the results reported that researchers were concerned about the generalizability of the participants' results and that participants would have poor reactions to receiving "bad" results [40].

Supplementary Articles

There were 140 supplementary articles (Multimedia Appendix 3), including the 8 articles identified in the primary search. The inclusion of these articles illustrates the return of results more broadly beyond digital systems.

In total, 98 articles reported genetics studies, with 27 relating specifically to *BRCA*1 and *BRCA2*. A total of 20 articles came

XSL•FO
RenderX

from environmental health, mostly related to chemical exposure, and 16 articles came from the general health domain. One study was related to staff retention, and another to family studies.

The return of aggregate results was described in 36 articles, the return of personalized results in 114 articles, and the return of nonidentifiable results in 3 articles. Additional education materials were included alongside results in 104 studies, while 11 articles did not report on this.

Of the possible methods for disseminating results either fully or partially (in cases where multiple methods were used), 56 studies delivered results via an in-person interaction, 48 used letters through traditional mail, 35 used phone calls, 25 used an online portal or application, 5 used emails for disseminating their results, 4 used other non-traditional methods, and in 32 articles the method used was unspecified. Digital methods, such as email and online portals, have become more common in recent years (Figure 2).

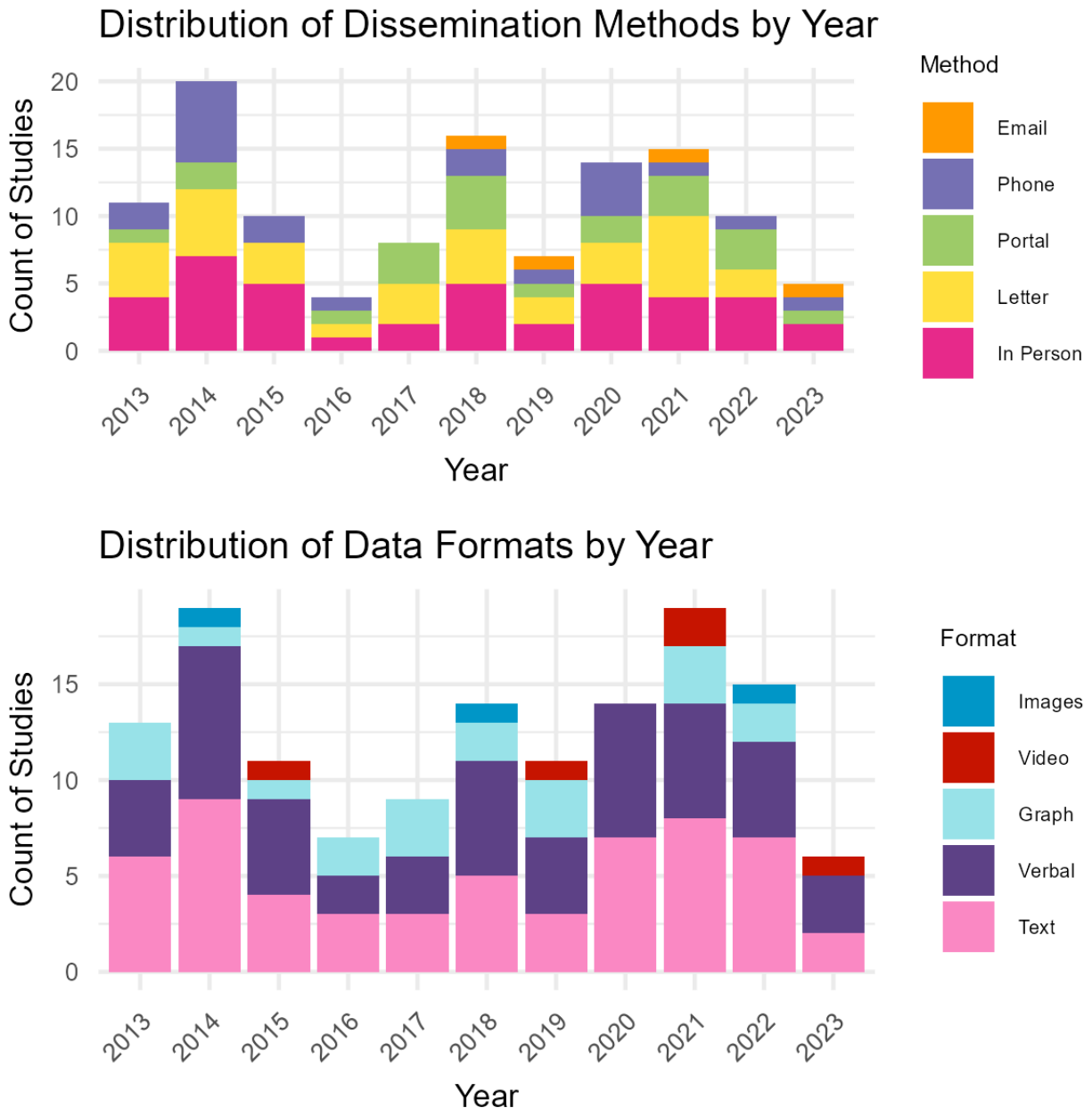
The data format also varied with many studies using multiple modalities: 85 studies returned results verbally, 73 in text format, and eight in video format (including YouTube videos or "relaxation tapes"). Some additional extraneous formats included t-shirts, folders, infographics, booklets, colour-in posters and comics, and various presentations, including conferences with an unspecified audience. More diverse methods have become more common in recent years (Figure 2).

In total, 29 studies included charts or graphs in their returned results (6 included graphics, images, or other visuals, including "visual aids," and 26 did not specify the format used).

A total of 85 studies involved experts to return results: 79 studies involved clinicians, with 62 including genetic counselors. Experts used to return results in other studies included a medical geneticist, an oncologist, physicians, nurses, and other HCPs. Other nonclinician experts included health educators, licensed social workers, laboratories, community health workers, community leaders, public health authorities, and decision-makers.

XSL•FO
RenderX

Figure 2. (Top) Distribution of dissemination methods of returned results (if reported) per publication year. (Bottom) Distribution of data formats of returned results (if reported) per publication year.



Discussion

Principal Findings

This review identified 8 articles that described and evaluated 7 systems for returning results to study participants and a total of 140 studies in which results were returned to participants, with or without using a specifically designed system. Among the 8 articles, 4 studies had dedicated digital platforms [35,37-39] to create results, share results, or both, while 139 of the total 140 studies used less scalable and costly methods, such as in-person meetings and phone calls, to return some or all results to participants.

Many researchers and research ethics boards are concerned about providing adequate participant privacy [27], and this concern is particularly relevant when participants receive personalized results. Typically, including personal information is required for the information to be valuable to the participant (such as receiving personal genetic results instead of study-level results), yet some researchers opted to preserve participant privacy at the cost of the relevancy of the results [43]. There may be a mismatch between the privacy risk participants perceive versus the privacy risk researchers believe participants perceive, as researchers tend to self-censor when considering research methods, including the return of results [37,44-46]. For example, while many researchers would likely consider t-shirts with individual data for participants [40] to be a blatant misuse of private data and put participants at risk, such personalized tokens were well received by the participants in that study as it aided in them forming a community, not only with the researchers but amongst themselves [40]. Such non-traditional options may be permissible by ethical research

XSL•FO
RenderX

frameworks but might not be considered due to researcher uncertainty around what is permitted and misconceptions of what may be beneficial to participants (such as the opportunity for peer support). While specific privacy and security measures are crucial, such as password-protected portals [38] or anonymizing data where possible [36], community benefits may be another factor to consider.

Dissemination modes constrain potential data formats. For example, returning graphical data over a phone call is impossible. The COVID-19 pandemic required a switch in modalities for all research fields [39,47-49], and more recent studies often required genetic counseling to be conducted over the phone rather than in person, where physical documents could be exchanged [36].

The accessibility of dissemination modalities should also be a consideration. Returning results in physical formats (such as physical documents, phone calls, or in-person meetings) often limits the ability to return results at scale due to resource limitations. However, transitioning to a fully digital system is not equitable for all participants for reasons including age, race, socioeconomic status, location, and language proficiency [50,51]. The ability to provide results in different formats facilitates the opportunity for participants to make use of the information for themselves [27,36,40]. This accessibility becomes even more crucial with participants from vulnerable populations, as they often have unique accessibility concerns [37], but also have increased incentives for getting information from research projects [37,38,40].

An additional consideration for creating accessible results is including adequate language and educational or contextual materials alongside results. While some participants may be well-versed in the research context due to previous research experience, most are laypeople. Therefore, considerations must be made regarding literacy and education [52]. Even when attempts were made to contextualize or explain complex concepts by providing additional educational materials, many participants still struggled to understand their results and what to do with the information [36-38]. While improving lay formats will always be valuable [53], additional prestudy or preresults educational materials may help bridge the gap.

Key Requirements of a Results-Sharing System

First, it is crucial to support multiple dissemination methods, such as supplying physical documents alongside digital text [27,36,38,40]. Digital systems, such as secure portals, may require considerable development effort but offer easy-to-use and scalable solutions [35,38,41]. However, additional physical formats like mailed documents should be available to participants as an option for the few who require them [27,40]. This ensures that participants who lack digital access or prefer nondigital interactions can still receive their results [40]. Participants who have different accessibility needs may be able to make use of existing digital adaptive technology [37] unless the researcher offers methods that support this intrinsically.

Secondly, text results may be adequate, though including relevant graphs, charts, and images can relay information more effectively [27,35,37,38,40,41,54]. Importantly, these must be in appropriate lay language, and educational materials must be provided to contextualize the results for participants [35,54,]. While participants are willing to receive favorable, neutral, and unfavorable results [37,38,40], excessive jargon or lack of clear directions of what can be done with the information impedes its use [36,38]. In many cases, participants will not act on neutral or positive information (such as clear screening results or standard contaminant levels) [38]. However, those who receive unfavorable information (such as indication of illness or other concern) express a clear desire to act on it and must be informed of who they can contact regarding their results, whether a research team member, their own personal HCP, or another expert for further information when results are unclear [36,38,40].

Thirdly, participants preferred receiving personalized information, while expecting many systems used methods to still maintain their privacy [35-37]. Articles emphasized the importance of deidentification whenever possible, which can involve using participant IDs or removing identifiers when storing and handling data up until the point of result dissemination, which can be further blinded from researchers by the system. This approach ensures that researchers do not have access to identifiable information during the result-creation process. Identifiers should only be used at the dissemination point to accurately link the results back to the participants, preserving their privacy and minimizing potential misuse of personalized data [35-37].

Strengths and Limitations

This review surveyed articles across various research disciplines using a registered protocol. A reference librarian aided in developing and validating a comprehensive search strategy and methodology to avoid bias toward a specific research discipline.

However, there are also several limitations. Returning results is rarely the primary focus of research articles and is a method, not an intervention. As there is no scalable method for screening the body text of published work, this review was limited to articles where the method of returning results was reported in the abstract or title of the article. Similarly, the return of results is referred to using different terminology between research disciplines, so it is possible that instances of returning results were missed as the specific terminology was not included in the search strategy; this may bias our results, due to the authors' greater familiarity with the terminology in some research fields rather than others. This may have also resulted in over-representing particular disciplines, primarily through iterative bibliography screening. In addition, this review did not consider any non-English language publications or publications after May 2023, which may have missed global or more contemporary research.

Future Work

Future research should explore returning results to healthy participants or less vulnerable groups to explore the influence of altruism and the benefit to all research participants rather than only those who have a clear benefit to participating.

A generalizable results-sharing system should have a discipline-independent design available to all researchers. Many

XSL•FO
RenderX

systems are highly tailored to a specific data and discipline type and, therefore, are unavailable to other researchers [27,35,37-40]. While this may contradict the requirement to provide educational and contextualizing materials, current solutions are too context-specific to be broadly adopted by other disciplines where participants could also benefit from receiving results. Consequently, this impedes their ability to scale and generalize, which can be addressed in future system design.

In addition, while most studies describe the methods used to return results, few include evaluation data to assess their effectiveness. Future research should incorporate evaluation components to determine whether these methods achieve their intended goals, such as increasing participant knowledge, reducing anxiety, or improving engagement with research findings. Moreover, none of the included studies compared different modalities for returning results, as they each focused on a single system design. Future work should explore participant preferences for different return methods to inform more user-centred approaches.

Conclusions

This review identified 7 systems used to return individual results to research participants. While the return of results systems and methods are often tailored to the needs of the participants and the research discipline, some factors are universal. There is a clear need to support multiple dissemination methods, such as providing results digitally via a portal and (if requested) in physical form to allow equal opportunity for participants to receive and share their results. The inclusion of additional educational materials is crucial. Text is the most standard data format for results, though other, more creative formats can be well received by participants when applicable. Future work should include the development and testing of a discipline-independent research results sharing system.

Acknowledgments

The authors thank Colleen Pawliuk (librarian at the University of British Columbia) for her feedback on the search strategy development and Alex Stolz, Jenna Ramji, and Sasha Sokolov for their help in screening articles for the review.

Generative artificial intelligence tools were not used in this manuscript's development, writing, or editing. The study was supported by an NSERC Discovery Grant (RGPIN-2021 - 02833 to MG). MG holds a Michael Smith Health Research BC scholar award (SCH-2020 - 0494).

Authors' Contributions

RC and MG were responsible for the conceptualization and methodology of the study. Investigation and data curation were conducted by RC and VC. The original draft was prepared by RC, and all authors critically reviewed the manuscript. Resources and supervision were provided by MG.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for the narrative review.
[DOCX File, 15 KB - i-jmr_v14i1e65606_app1.docx]

Multimedia Appendix 2
Data extraction form for narrative review.
[DOCX File, 17 KB - i-jmr_v14i1e65606_app2.docx]

Multimedia Appendix 3
Table of Supplementary Articles included in the narrative review.
[DOCX File, 200 KB - i-jmr_v14i1e65606_app3.docx]

Multimedia Appendix 4
Risk of Bias analysis.
[DOCX File, 19 KB - i-jmr_v14i1e65606_app4.docx]

Checklist 1
PRISMA Checklist
[PDF File, 528 KB - i-jmr_v14i1e65606_app5.pdf]

References

XSL•FO
RenderX

1. Kumar D, Shen K, Case B, et al. All things considered: an analysis of iot devices on home networks. Presented at: URL: https://www.usenix.org/conference/usenixsecurity19/presentation/kumar-deepak [accessed 2025-07-28]
2. Pew research center. Mobile fact sheet Pew Research Center fact sheets: Tech adoption trends. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2023-04-03]
3. Bonney R, Shirk JL, Phillips TB, et al. Next steps for citizen science. Science 2014 Mar 28;343(6178):1436-1437. [doi: 10.1126/science.1251554]
4. Jollymore A, Haines MJ, Satterfield T, Johnson MS. Citizen science for water quality monitoring: data implications of citizen perspectives. J Environ Manage 2017 Sep 15;200:456-467. [doi: 10.1016/j.jenvman.2017.05.083] [Medline: 28618317]
5. Raddick MJ, Bracey G, Gay PL, et al. Galaxy zoo: exploring the motivations of citizen science volunteers. Astronomy Education Review 2010 Dec;9(1). [doi: 10.3847/AER2009036]
6. Conrad CC, Hilchey KG. A review of citizen science and community-based environmental monitoring: issues and opportunities. Environ Monit Assess 2011 May;176(1-4):273-291. [doi: 10.1007/s10661-010-1582-5] [Medline: 20640506]
7. Lai AM, Hsueh PY, Choi YK, Austin RR. Present and future trends in consumer health informatics and patient-generated health data. Yearb Med Inform 2017 Aug;26(1):152-159. [doi: 10.15265/IY-2017-016] [Medline: 29063559]
8. Omoloja A, Vundavalli S. Patient generated health data: benefits and challenges. Curr Probl Pediatr Adolesc Health Care 2021 Nov;51(11):101103. [doi: 10.1016/j.cppeds.2021.101103] [Medline: 34799255]
9. Lordon RJ, Mikles SP, Kneale L, et al. How patient-generated health data and patient-reported outcomes affect patient-clinician relationships: a systematic review. Health Informatics J 2020 Dec;26(4):2689-2706. [doi: 10.1177/1460458220928184] [Medline: 32567460]
10. Coppola L, Cianflone A, Grimaldi AM, et al. Biobanking in health care: evolution and future directions. J Transl Med 2019 May 22;17(1):172. [doi: 10.1186/s12967-019-1922-3] [Medline: 31118074]
11. Seltzer E, Goldshear J, Guntuku SC, et al. Patients' willingness to share digital health and non-health data for research: a cross-sectional study. BMC Med Inform Decis Mak 2019 Aug 8;19(1):157. [doi: 10.1186/s12911-019-0886-9] [Medline: 31395102]
12. Maus B, Salvi D, Olsson CM, 10th Int Conf Internet Things Companion. : Association for Computing Machinery; 2020 Presented at: Enhancing Citizens Trust in Technologies for Data Donation in Clinical Research: Validation of a Design Prototype p. 1-8. [doi: 10.1145/3423423.3423430]
13. Huchuk B, O'Brien W, Sanner S. A longitudinal study of thermostat behaviors based on climate, seasonal, and energy price considerations using connected thermostat data. Build Environ 2018 Jul;139:199-210. [doi: 10.1016/j.buildenv.2018.05.003]
14. Snider C. Tidepool. Let's talk about your donated data (mg/dL). URL: https://www.tidepool.org/blog/lets-talk-about-your-donated-data [accessed 2022-09-27]
15. Skatova A, Goulding J. Psychology of personal data donation. PLOS ONE 2019;14(11):e0224240. [doi: 10.1371/journal.pone.0224240] [Medline: 31747408]
16. Holl F, Reschke L, Muller-Miellitz S, Fotteler M, Swoboda W. Secondary use of research data: a review of availability and utilization of public use files and initial steps for the development of A process to facilitate medical research data donation. 2019 Presented at: 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA) p. 1-5. [doi: 10.1109/AICCSA47632.2019.9035331]
17. Holl F, Fotteler M, Müller-Mielitz S, Swoboda W. Investigating ethical, legal, and socio-technical barriers of medical data donation and developing a concept to address them - a research protocol. Stud Health Technol Inform 2019 Jul 4;262:51-54. [doi: 10.3233/SHTI190014] [Medline: 31349263]
18. Baines R, Stevens S, Austin D, et al. Patient and public willingness to share personal health data for third-party or secondary uses: systematic review. J Med Internet Res 2024 Mar 5;26:e50421. [doi: 10.2196/50421] [Medline: 38441944]
19. Committee on the Return of Individual-Specific Research Results Generated in Research Laboratories, Board on Health Sciences Policy, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. In: Botkin JR, Mancher M, Busta ER, Downey AS, editors. Returning Individual Research Results to Participants: Guidance for a New Research Paradigm: National Academies Press; 2018:25094. [doi: 10.17226/25094]
20. Sobel ME, Dreyfus JC, Dillehay McKillip K, et al. Return of Individual Research Results: A Guide for Biomedical Researchers Utilizing Human Biospecimens. Am J Pathol 2020 May;190(5):918-933. [doi: 10.1016/j.ajpath.2020.01.014] [Medline: 32201265]
21. Mash C, McAllister KA, Wonnum S, et al. Principles and practices of returning individual research results to participants in large studies of pregnancy and childhood. Am J Epidemiol 2025 Mar 4;194(3):830-836. [doi: 10.1093/aje/kwae228] [Medline: 39030726]
22. Vears DF, Hallowell N, Bentzen HB, et al. A practical checklist for return of results from genomic research in the European context. Eur J Hum Genet 2023 Jun;31(6):687-695 [FREE Full text] [doi: 10.1038/s41431-023-01328-6] [Medline: 36949262]
23. Lázaro-Muñoz G, Torgerson L, Smith HS, Pereira S. Perceptions of best practices for return of results in an international survey of psychiatric genetics researchers. Eur J Hum Genet 2021 Feb;29(2):231-240. [doi: 10.1038/s41431-020-00738-0] [Medline: 33011736]

XSL•FO
RenderX

24. Long CR, Stewart MK, Cunningham TV, Warmack TS, McElfish PA. Health research participants' preferences for receiving research results. Clin Trials 2016 Dec;13(6):582-591. [doi: 10.1177/1740774516665598] [Medline: 27562368]
25. Melvin CL, Harvey J, Pittman T, Gentilin S, Burshell D, Kelechi T. Communicating and disseminating research findings to study participants: Formative assessment of participant and researcher expectations and preferences. J Clin Transl Sci 2020 Jan 20;4(3):233-242. [doi: 10.1017/cts.2020.9] [Medline: 32695495]
26. Richmond J, Cunningham-Erves J, Givens B, et al. All of us participant perspectives on the return of value in research. Genet Med 2024 Aug;26(8):101163. [doi: 10.1016/j.gim.2024.101163] [Medline: 38738530]
27. Boronow KE, Susmann HP, Gajos KZ, et al. DERBI: a digital method to help researchers offer "right-to-know" personal exposure results. Environ Health Perspect 2017 Feb 1;125(2):A27-A33. [doi: 10.1289/EHP702] [Medline: 28145870]
28. Sandhaus S, Kaufmann D, Ramirez-Andreotta M. Public participation, trust and data sharing: gardens as hubs for citizen science and environmental health literacy efforts. Int J Sci Educ B Commun Public Engagem 2019;9(1):54-71. [doi: 10.1080/21548455.2018.1542752] [Medline: 31485378]
29. Lebow-Skelley E, Yelton S, Janssen B, Erdei E, Pearson MA. Identifying issues and priorities in reporting back environmental health data. Int J Environ Res Public Health 2020 Sep 16;17(18):32947900. [doi: 10.3390/ijerph17186742] [Medline: 32947900]
30. Majdoubi A, O'Connell SE, Michalski C, et al. Antibody reactivity against SARS-cov-2 in adults from the vancouver metropolitan area, canada. medRxiv. Preprint posted online on Nov 6, 2020. [doi: 10.1101/2020.10.05.20206664]
31. Carr R, West NC, Görges M. Development of a Standardized Set of Requirements for Systems to Return Individual Research Results Back to Study Participants: Narrative Review: Open Science Foundation (OSF) Registries; 2023. [doi: 10.17605/OSF.IO/RD3PV]
32. Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? Eur J Clin Invest 2018 Jun;48(6):e12931. [doi: 10.1111/eci.12931] [Medline: 29578574]
33. Tyndall J. AACODS Checklist for appraising grey literature.: Flinders University; 2010. URL: https://fac.flinders.edu.au/dspace/api/core/bitstreams/e94a96eb-0334-4300-8880-c836d4d9a676/content [accessed 2025-07-23]
34. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
35. Savatt JM, Azzariti DR, Faucett WA, et al. ClinGen's GenomeConnect registry enables patient-centered data sharing. Hum Mutat 2018 Nov;39(11):1668-1676. [doi: 10.1002/humu.23633] [Medline: 30311371]
36. Ohneda K, Hiratsuka M, Kawame H, et al. A pilot study for return of individual pharmacogenomic results to population-based cohort study participants. JMA J 2022;5(2):177-189. [doi: 10.31662/jmaj.2021-0156]
37. Gilbert RM, Sumodhee D, Pontikos N, et al. Collaborative research and development of a novel, patient-centered digital platform (MyEyeSite) for rare inherited retinal disease data: acceptability and feasibility study. JMIR Form Res 2022 Jan 31;6(1):e21341. [doi: 10.2196/21341] [Medline: 35099396]
38. Cope H, Lincoln-Boyea B, Gwaltney AY, et al. Use of a web-based portal to return normal individual research results in early check: exploring user behaviors and attitudes. Clin Genet 2023 Jun;103(6):672-680. [doi: 10.1111/cge.14325] [Medline: 36960544]
39. van de Poll-Franse LV, Horevoorts N, Schoormans D, et al. Measuring clinical, biological, and behavioral variables to elucidate trajectories of patient-reported outcomes: the PROFILES registry. J Natl Cancer Inst 2022 Jun 13;114(6):800-807. [doi: 10.1093/jnci/djac047] [Medline: 35201353]
40. Perovich LJ, Ohayon JL, Cousins EM, et al. Reporting to parents on children's exposures to asthma triggers in low-income and public housing, an interview-based case study of ethics, environmental literacy, individual action, and public health benefits. Environ Health 2018 May 21;17(1):48. [doi: 10.1186/s12940-018-0395-9] [Medline: 29784007]
41. Polka E, Childs E, Friedman A, et al. MCR: Open-source software to automate compilation of health study report-back. Int J Environ Res Public Health 2021 Jun 5;18(11):6104. [doi: 10.3390/ijerph18116104] [Medline: 34198866]
42. Dragicevic P, Jansen Y, Vande Moere A. Data physicalization. In: Vanderdonckt J, Palanque P, Winckler M, editors. Handb Hum Comput Interact: Springer International Publishing; 2021:1-51. [doi: 10.1007/978-3-319-27648-9_94-1]
43. Pulford DJ, Harter P, Floquet A, et al. Communicating BRCA research results to patients enrolled in international clinical trials: lessons learnt from the AGO-OVAR 16 study. BMC Med Ethics 2016 Dec;17(1):63. [doi: 10.1186/s12910-016-0144-y]
44. Beil A, Hornsby W, Uhlmann WR, et al. Disclosure of clinically actionable genetic variants to thoracic aortic dissection biobank participants. BMC Med Genomics 2021 Dec;14(1):66. [doi: 10.1186/s12920-021-00902-5]
45. Keller MA, Gordon ES, Stack CB, et al. Coriell personalized medicine collaborative®: a prospective study of the utility of personalized medicine. Per Med 2010 May;7(3):301-317. [doi: 10.2217/pme.10.13] [Medline: 29776223]
46. Bernaerdt J, Moerenhout T, Devisch I. Vulnerable patients' attitudes towards sharing medical data and granular control in patient portal systems: an interview study. J Eval Clin Pract 2021 Apr;27(2):429-437. [doi: 10.1111/jep.13465] [Medline: 32886953]
47. Best LG, O'Leary M, O'Leary R, Lawrence W, Torgerson DG. Return of participants' incidental genetic research findings: experience from a case-control study of asthma in an American Indian community. Res Sq. 2023 Jan 19 p. rs.3.rs-2485539. [doi: 10.21203/rs.3.rs-2485539/v1] [Medline: 36711525]

XSL•FO
RenderX

48. Muse ED, Chen SF, Liu S, et al. Impact of polygenic risk communication: an observational mobile application-based coronary artery disease study. NPJ Digit Med 2022 Mar 11;5(1):30. [doi: 10.1038/s41746-022-00578-w] [Medline: 35277577]
49. Salvi N, Itta KC, Lachyan A, et al. Experiences of sharing results of community based serosurvey with participants in a district of Maharashtra, India. PLOS ONE 2022;17(8):e0271920. [doi: 10.1371/journal.pone.0271920] [Medline: 35925959]
50. Grossman LV, Masterson Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 1;26(8-9):855-870. [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
51. Rodriguez JA, Shachar C, Bates DW. Digital inclusion as health care — supporting health care equity with digital-infrastructure initiatives. N Engl J Med 2022 Mar 24;386(12):1101-1103. [doi: 10.1056/NEJMp2115646]
52. Wilkins CH, Mapes BM, Jerome RN, Villalta-Gil V, Pulley JM, Harris PA. Understanding what information is valued by research participants, and why. Health Aff (Millwood) 2019 Mar;38(3):399-407. [doi: 10.1377/hlthaff.2018.05046]
53. George MS, Gaitonde R, Davey R, Mohanty I, Upton P. Engaging participants with research findings: A rights-informed approach. Health Expect 2023 Apr;26(2):765-773. [doi: 10.1111/hex.13701] [Medline: 36647684]
54. Ohneda K, Hamanaka Y, Kawame H, et al. Returning individual genomic results to population-based cohort study participants with BRCA1/2 pathogenic variants. BREAST CANCER 2023 Jan;30(1):110-120. [doi: 10.1007/s12282-022-01404-7] [Medline: 36161580]

Abbreviations

HCP: health care provider
PRISMA: Preferred Reporting Items for Systematic reviews and Meta-Analyses

Edited by TDA Cardoso; submitted 20.08.24; peer-reviewed by H Cope, LK Chua; revised version received 01.05.25; accepted 05.05.25; published 18.08.25.

Please cite as:
Carr RL, Chan V, West NC, Görges M
Developing Requirements for a Standardized System to Return Individual Research Results Back to Study Participants: Narrative Review
Interact J Med Res 2025;14:e65606
URL: https://www.i-jmr.org/2025/1/e65606
doi:10.2196/65606

© Rosalyn Leigh Carr, Vita Chan, Nicholas C West, Matthias Görges. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.8.2025. 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

Economic Evaluations and Equity in the Use of Artificial Intelligence in Imaging Examinations for Medical Diagnosis in People With Dermatological, Neurological, and Pulmonary Diseases: Systematic Review

Giulia Osório Santana[1,2], BS; Rodrigo de Macedo Couto[2], PhD; Rafael Maffei Loureiro[2], MD; Brunna Carolinne Rocha Silva Furriel[2,3,4], MSc; Luis Gustavo Nascimento de Paula[1], PhD; Edna Terezinha Rother[5], BS; Joselisa Péres Queiroz de Paiva[2], PhD; Lucas Reis Correia[1,6], MSc

[1]PROADI-SUS, Hospital Israelita Albert Einstein, 462 Madre Cabrini Street, Tower A, 5th Floor, São Paulo, SP, Brazil

[2]Department de Imagem, Hospital Israelita Albert Einstein, São Paulo, Brazil

[3]Escola de Engenharia de Computação, Universidade Federal de Goiás, Goiânia, Brazil

[4]Grupo de Estudos e Pesquisa em Ciência e Tecnologia (GCITE), Instituto Federal de Goiás, Goiânia, Brazil

[5]Instituto Israelita de Ensino e Pesquisa, Hospital Israelita Albert Einstein, São Paulo, Brazil

[6]Departamento de Medicina Preventiva, Universidade de São Paulo, São Paulo, Brazil

Corresponding Author:
Giulia Osório Santana, BS
PROADI-SUS, Hospital Israelita Albert Einstein, 462 Madre Cabrini Street, Tower A, 5th Floor, São Paulo, SP, Brazil

Abstract

Background: Health care systems around the world face numerous challenges. Recent advances in artificial intelligence (AI) have offered promising solutions, particularly in diagnostic imaging.

Objective: This systematic review focused on evaluating the economic feasibility of AI in real-world diagnostic imaging scenarios, specifically for dermatological, neurological, and pulmonary diseases. The central question was whether the use of AI in these diagnostic assessments improves economic outcomes and promotes equity in health care systems.

Methods: This systematic review has 2 main components, economic evaluation and equity assessment. We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) tool to ensure adherence to best practices in systematic reviews. The protocol was registered with PROSPERO (International Prospective Register of Systematic Reviews), and we followed the PRISMA-E (Preferred Reporting Items for Systematic Reviews and Meta-Analyses - Equity Extension) guidelines for equity. Scientific articles reporting on economic evaluations or equity considerations related to the use of AI-based tools in diagnostic imaging in dermatology, neurology, or pulmonology were included in the study. The search was conducted in the PubMed, Embase, Scopus, and Web of Science databases. Methodological quality was assessed using the following checklists, CHEC (Consensus on Health Economic Criteria) for economic evaluations, EPHPP (Effective Public Health Practice Project) for equity evaluation studies, and Welte for transferability.

Results: The systematic review identified 9 publications within the scope of the research question, with sample sizes ranging from 122 to over 1.3 million participants. The majority of studies addressed economic evaluation (88.9%), with most studies addressing pulmonary diseases (n=6; 66.6%), followed by neurological diseases (n=2; 22.3%), and only 1 (11.1%) study addressing dermatological diseases. These studies had an average quality access of 87.5% on the CHEC checklist. Only 2 studies were found to be transferable to Brazil and other countries with a similar health context. The economic evaluation revealed that 87.5% of studies highlighted the benefits of using AI in dermatology, neurology, and pulmonology, highlighting significant cost-effectiveness outcomes, with the most advantageous being a negative cost-effectiveness ratio of –US $27,580 per QALY (quality-adjusted life year) for melanoma diagnosis, indicating substantial cost savings in this scenario. The only study assessing equity, based on 129,819 radiographic images, identified AI-assisted underdiagnosis, particularly in certain subgroups defined by gender, ethnicity, and socioeconomic status.

Conclusions: This review underscores the importance of transparency in the description of AI tools and the representativeness of population subgroups to mitigate health disparities. As AI is rapidly being integrated into health care, detailed assessments are essential to ensure that benefits reach all patients, regardless of sociodemographic factors.

Trial Registration: PROSPERO CRD42023407755; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023407755

International Registered Report Identifier (IRRID): RR2-10.2196/48544

XSL•FO
RenderX

(Interact J Med Res 2025;14:e56240) doi:10.2196/56240

KEYWORDS

machine learning; ML; artificial intelligence; AI; algorithm; algorithms; economic evaluation; equity

Introduction

Background

Health care systems globally face numerous challenges, including geographical and economic disparities, uneven access to services, and unequal distribution of health care professionals. Patient diversity, influenced by factors such as skin color, ethnicity, and socioeconomic status, adds an extra layer of complexity [1]. However, over the last several years, technological advances in artificial intelligence (AI) have offered promising solutions [2,3]. AI, an interdisciplinary field, aims to develop systems capable of performing tasks that typically require human reasoning [4]. Machine learning, a subfield of AI, creates algorithms capable of learning and improving from experience without explicit programming [5,6]. In health care, AI algorithms have been developed to improve medical image analysis, offering benefits such as automatic recognition of abnormalities and disease prediction, ultimately leading to optimized clinical decision-making, cost reduction, and error prevention [7,8]. Despite these advances, gaps remain in evaluating the economic feasibility and equity implications of AI tools in real-world applications, as highlighted by previous studies [9]. Some critical issues must be addressed to maximize the clinical potential of AI. Methodological limitations, including insufficient validation studies and the lack of transparency in the AI decision-making process, threaten the broader adoption and effectiveness of AI systems in clinical settings [10,11].

Moreover, an emerging issue is bias in the real-world use of AI algorithms, particularly with respect to the populations used to train and validate AI models, potentially leading to skewed results that favor certain ethnic or demographic groups [12]. This bias can result in less effective or misleading results when applied to underrepresented populations, perpetuating existing health care inequalities [13,14]. Voets et al [9] emphasized these disparities, noting the critical need for research that not only evaluates the cost-effectiveness of AI but also addresses how these technologies can reduce inequities in health care access.

Equity in health care advocates for equal access to quality care regardless of personal characteristics such as ethnicity, gender, or socioeconomic status [15]. The inclusion of equity-related data into AI requires careful interpretation to avoid unintended biases [16-19]. Though the relationship between economic evaluation and equity lies in the distributional impact of costs and benefits. By identifying cost-effective strategies, economic evaluations can help allocate resources more equitably, ensuring that underserved populations gain access to innovative diagnostic technologies. For instance, reducing diagnostic costs through AI could enhance access for low-income or remote populations, addressing disparities in health care delivery[18].

Brazil, an upper-middle–income country with many social inequalities, has the largest free public health care system in the world, on which the majority of its population relies [20]. However, socioeconomic barriers create challenges in providing diagnoses, especially in impoverished and remote areas. Digital health initiatives, including AI-powered clinical decision support systems, have proven effective in reducing health disparities and improving access to health services [21]. To tackle these challenges, the Brazilian Ministry of Health's Support Program for the Institutional Development of the Unified Health System (in Portuguese, "Programa de Apoio ao Desenvolvimento Institucional do Sistema Único de Saúde," known as PROADI-SUS) supports a collaborative project named "Banco de Imagens" (Bank of Images) to create a nationwide cloud-based repository of medical images, as well as to develop and validate AI algorithms to assist in disease diagnosis.

The selection of dermatological, neurological, and pulmonary diseases was guided by the Brazilian Ministry of Health, which sponsors this study, as these conditions are considered national health priorities. Dermatological diseases, particularly melanoma, were included due to the critical need for early diagnosis and the shortage of specialists. Neurological diseases rely heavily on imaging for diagnosis, where timely intervention is essential; yet, specialist availability remains limited. Pulmonary diseases, particularly tuberculosis (TB), continue to pose significant public health challenges, especially in low-resource settings. The Ministry of Health identified these 3 disease areas as key targets where AI could enhance screening and diagnosis. Given this strategic priority, we conducted a single study encompassing all 3 disease groups.

Objectives

This systematic review investigates the economic performance and viability of the use of AI in diagnostic imaging examinations in real-world scenarios, particularly for dermatological, neurological, and pulmonary diseases. This review will also assess the potential impact of this intervention on equity in health care systems. The research question guiding this investigation is as follows: "To what extent does the use of AI in imaging examinations for the diagnosis of dermatological, neurological, and pulmonary diseases result in improved economic outcomes, and does it promote equity in health care systems?"

Methods

Overview

This study was conducted as a systematic review with two main components. The first component focused on economic evaluations, while the second focused on equity considerations. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) tool [22] was used to ensure adherence to best practices in systematic reviews (the PRISMA checklist is provided in Checklist 1). A protocol for this systematic review was registered in the PROSPERO (International Prospective Register of Systematic Reviews)

XSL•FO
RenderX

database under registration number CRD42023407755. No deviations from the registered protocol occurred during the study. We have published a detailed version of the protocol [23]. In addition, we followed the PRISMA-E (Preferred Reporting Items for Systematic Reviews and Meta-Analyses - Equity Extension) guidelines for systematic reviews with a focus on equity [24]. To structure our research question, we followed the PICO (Population, Intervention, Comparison, and Outcome) framework, as shown in Table 1.

Table . Population, Intervention, Comparison, and Outcome structure for the systematic review of global economic evaluations and equity studies on the use of artificial intelligence in medical diagnosis in people with dermatological, neurological, and pulmonary diseases.

Acronym	Description	Criteria
P	Population	Patients examined for pulmonary, neurological, and dermatological diseases, considering the entire world population, regardless of income and development of the respective country
I	Intervention	Use of artificial intelligence to support diagnostic decisions in any type of medical imaging
C	Comparison	Conventional or human-based diagnostic methods
O	Outcomes	Full economic evaluation studies, such as cost-effectiveness, cost-utility, cost-benefit, or cost-minimization. We also assessed changes in patient outcomes related to health care access and demographic factors such as age, gender, race, and income, with a focus on equity.

Eligibility Criteria

Scientific articles reporting on economic evaluations or equity considerations related to the use of AI-based tools in diagnostic imaging. These include (1) dermatology, specifically melanoma and carcinoma; (2) neurology, focusing on conditions with notable radiological findings such as microcephaly, brain atrophy, and hydrocephalus; and (3) pulmonology, focusing on conditions such as TB, lung consolidation, pleural effusion, atelectasis, pneumothorax, mediastinal widening, lung edema, lung opacity, lung lesion, lung cancer, other pleural diseases, and lung nodules. Detailed inclusion and exclusion criteria, including language and year of publication, can be found in Table 2.

Table . Inclusion and exclusion criteria for the systematic review of global economic evaluations and equity studies on the use of artificial intelligence (AI) in medical diagnosis in people with dermatological, neurological, and pulmonary diseases.

Criteria	Inclusion	Exclusion
Article types	Original articles reporting economic or equity evaluations. Cost-effectiveness, cost-utility, cost-benefit, and cost-minimization analyses were considered. Regarding equity, the authors analyzed reported improvements or declines in patient outcomes based on access to health care and demographic indicators such as age, gender, race, and income.	Articles that were not complete economic evaluations, such as cost analyses, cost description studies, and cost outcome descriptions, were excluded. The exclusion was based on the need to include only studies that provided a complete assessment of both costs and outcomes, allowing for a comprehensive understanding of the economic impact of AI interventions. This approach ensures that the review addresses studies that evaluate both the financial and health-related consequences of AI technologies. Regarding equity, we excluded studies that did not account for diverse samples reflective of the broader population.
Study focus	The use of artificial intelligence-based tools in diagnostic imaging in dermatological, pulmonary, and neurological fields.	Studies focused on other medical areas, such as heart disease, breast cancer, and diabetes. AI clinical validation and studies to predict risks were also excluded.
Language	No restrictions based on language.	No restrictions based on language.
Year of publication	No restrictions based on the year of publication	No restrictions based on the year of publication

Information Sources

Searches were conducted in the PubMed, Embase, Scopus, and Web of Science databases in March 2023, with no restrictions on publication date. Additional searches were performed to identify the references cited in the retrieved articles.

XSL•FO
RenderX

Search Strategy

A customized search strategy was defined for each database, using terms related to AI, relevant health conditions or diseases according to eligibility criteria, economic evaluations, and equity. A separate search was conducted for each topic area, as well as a general search without a specific area. The final search terms are provided in the Multimedia Appendices 1-4.

Data Management

An Excel (Microsoft Corp) spreadsheet was used to manage and organize the records during data extraction and synthesis.

Selection Process

First, all records retrieved by the search terms in different databases were exported to EndNote software (version X9.3.3; Clarivate), a reference management software. Then, the articles were exported to the Rayyan platform (Qatar Computing Research Institute), which was used to remove duplicates and select articles for screening by reading titles and abstracts.

The selection process consisted of 2 phases. In the first phase, 2 researchers independently assessed and selected articles. Researchers GOS and RML evaluated titles and abstracts from the fields of pulmonology and neurology, while researchers GOS and BCRS evaluated titles and abstracts from the field of dermatology. Researchers GOS and RMC also assessed an additional search related to the use of AI in medical images in general, without a specific field. All disagreements were resolved by a third researcher (LRC).

In the second phase, the selected articles from the first phase underwent full-text review. Researchers GOS and RMC assessed the full texts for all fields and searches. All disagreements were resolved by a third researcher (LRC).

Data Extraction Process

After selecting the articles, 2 researchers (GOS and RMC) collected the variables of interest for the systematic review from each publication with available data. The following information was prioritized for collection: publication title, year of publication, journal, authors, URL, study type, language, study location, time horizon, corresponding author contact, summary, objectives, methodology, study population, participant demographics, health indicators, access to health care services, equity indicators such as odds ratios and relative risks, study perspective, sample size, statistical significance, applied AI, incremental cost-effectiveness ratio (ICER), cost-benefit or equity of results, quality-adjusted life years (QALY), direction of effect (favorable or not), model applied, sensitivity analysis, initial assessment of methodological quality, conflicts of interest, funding, and transferability of results.

We adopted the World Bank's 2023 classification of countries by income level [25], which categorizes groups by gross national income per capita calculated using the World Bank's Atlas method. The groups are low income, US $1135 or less; lower-middle income, US $1136 - $4465; upper-middle income, US $4466 - $13,845; high income, US $13,845 or more.

Risk of Bias in Individual Studies and Transferability

Three tools were used to assess the methodological quality of the selected studies: (1) the CHEC (Consensus on Health Economic Criteria Checklist) economic evaluation checklist [26]; (i2) the EPHPP (Effective Public Health Practice Project) quality assessment tool for quantitative studies that are not economic evaluations but were selected for their equity results [27]; and (3) the Welte checklist to assess the transferability of technologies used in the selected studies between different countries [26,28]. The methodological quality assessment of the studies was conducted independently by 2 researchers (GOS and RMC). Any discrepancies were resolved through discussion with a third researcher (LGNP).

For the CHEC checklist, points were awarded for each criterion met, so that a maximum of 20 points were available for the evaluation. No points were awarded if the criterion was not fully met. A percentage score was calculated by dividing the number of points awarded by the total number of points on the checklist and multiplying by 100.

For the EPHPP tool, each section was scored 1 (strong) if it met the criteria, 2 (moderate) if at least 1 section was rated as weak, and 3 (weak) if 2 or more sections were rated as weak. Finally, a global score was calculated based on the scores for each section.

For the Welte checklist, points were assigned based on the following criteria: perspective (2 points); discount rate (2 points); medical cost approach (2 points); absolute and relative prices in health care (1 point); practice variation (1 point); incidence and prevalence (1 point); case mix (1 point); life expectancy (1 point); productivity and lost work time (1 point), for a total of 12 points. Studies that scored 10 points or more were considered transferable.

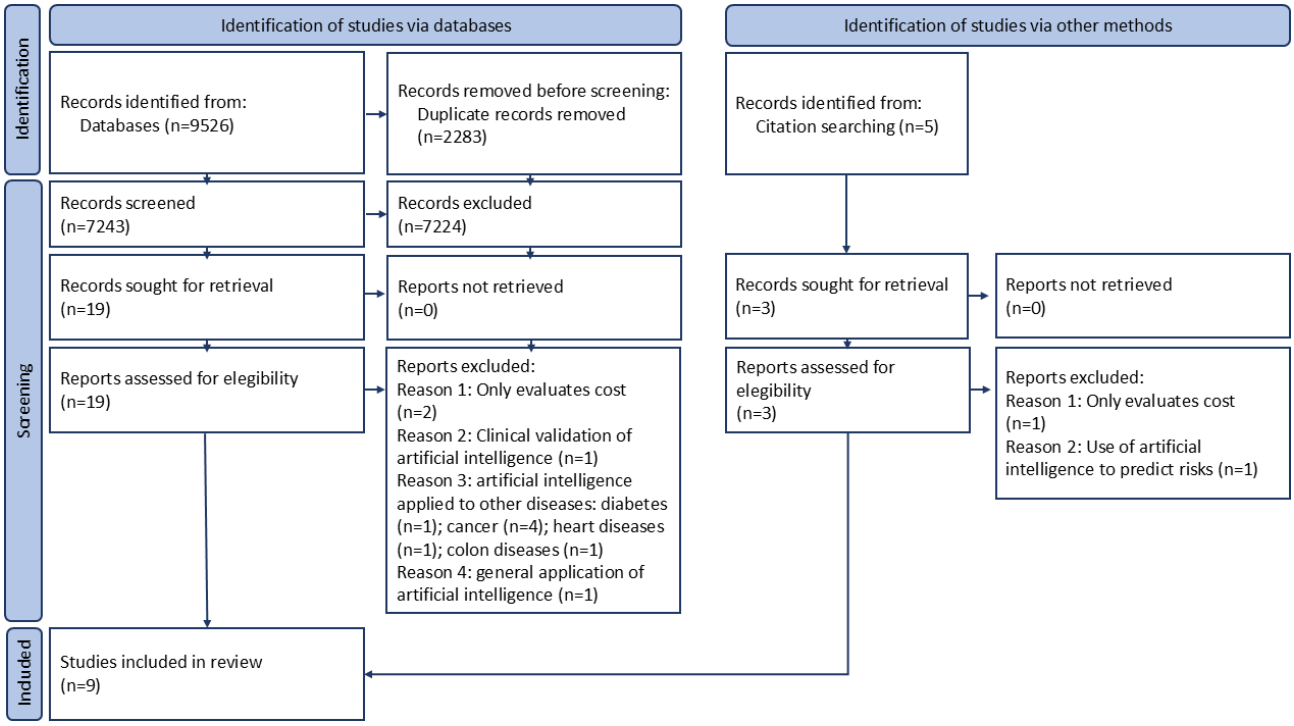
Results

Search Results

The database searches identified a total of 9526 entries, of which 2283 (23.9%) were excluded as duplicates. Of the remaining sample of 7243 articles, 7224 (99.7%) were excluded after screening of titles and abstracts. A total of 19 articles proceeded to the eligibility stage, of which 11 (57.9%) were excluded as they did not meet the inclusion and exclusion criteria. In addition, by reading the references of the selected articles, 3 further articles were identified for full-text reading, of which only 1 (33.3%) was selected for the study. A schematic representation delineating the process of record identification, screening, and exclusion is depicted in Figure 1.

XSL•FO
RenderX

Figure 1. Flowchart illustrating the selection of articles included in the study, based on database searches and additional sources.



Overview of the Included Studies

An overview of the included studies is presented in Table 3 and Multimedia Appendix 5. Most of these studies were published in 2021 (n=5, 55.5%), the others in 2022 (n=3, 33.4%) and 2020 (n=1, 11.1%). They focused primarily on pulmonary diseases (n=6, 66.6%; [29-34]), followed by neurological diseases (n=2, 22.3%; [35,36]), and only 1 [37] study (11.1%) focused on dermatological diseases. Of the selected studies, 8 [29-32,34-37] (88.9%) were economic evaluation studies and only 1 (11.1%) [33] addressed equity. The majority of studies (n=5, 55.5%) [29,33-35,37] were conducted in high-income countries, and 3 [30-32] of the 4 (75.0%) studies were conducted in low- to middle-income countries in partnership with institutions located in high-income countries. Notably, 5 [29,31,32,35,37] out of 9 studies (55.5%) reported a conflict of interest. Only 3 [31,32,37] out of 9 studies (33.3%) reported funding, none of which was private.

XSL•FO
RenderX

Table . General characteristics of the articles selected for the study and the results of the checklists applied for quality assessment.

ID	First author and year of publication	Study type	Place of study	Conflict of interest	Funding
Dermatology, Dentistry, and Ophthalmology					
1	Gomez Rossi et al [37]	Cost-effectiveness analysis	United States, Germany, and Brazil	Yes	Yes, public
Neurology					
2	Leeuwen et al [35]	Cost-effectiveness analysis	United Kingdom	Yes	No
3	Mansour et al [36]	Cost-effectiveness analysis	Egypt	No	Not mentioned
Pulmonology					
4	Adams et al [29]	Cost-minimization analysis	United States	Yes	No
5	Bashir et al [30]	Cost-minimization analysis	Pakistan	No	No
6	MacPherson et al [31]	Cost-effectiveness analysis	Malawi	Yes	Yes, philanthropic institution
7	Nsengiyumva et al [32]	Cost-effectiveness analysis	Pakistan	Yes	Yes, public
8	Seyyed-Kalantari et al [33]	Equity	United States	No	No
9	Ziegelmayer et al [34]	Cost-effectiveness analysis	United States	No	No

Methodological Quality Assessment

The methodological quality of the studies was assessed using the CHEC checklist, with scores calculated based on the percentage of criteria met (Tables 4 and 5). Scores ranged from 55.0% to 100.0%, with an average of 85.6%.

Table . Scores for the checklists applied for quality assessment.

ID	CHEC (Consensus Health Economic Criteria) score, (%)	EPHPP (Effective Public Health Practice Project) score	Welte score
1	95%	N/A[a]	10 points (transferable)
2	90%	N/A	10 points (transferable)
3	55%	N/A	6 points (nontransferable)
4	80%	N/A	9 points (nontransferable)
5	75%	N/A	8 points (nontransferable)
6	100%	N/A	7 points (nontransferable)
7	95%	N/A	8 points (nontransferable)
8	N/A	Strong (no weak sections)	N/A
9	95%	N/A	8 points (nontransferable)

[a]N/A: not applicable

XSL•FO
RenderX

Table . Results of the checklists applied for quality assessment.

ID	Replaced clinical practice	Description and use of AI[a] application	Study population	Sample size
1	Diagnosis of melanoma (dermatology) and diabetic retinopathy (ophthalmology) performed solely by clinicians without AI assistance	Convolutional Neural Network (CNN)[b]. Use: decision support for the diagnosis of melanoma, caries, and diabetes	General population (age 50 y), children (age 12 y), people with diabetes (over 40 y), respectively	N/A[c]
2	Detection of intracranial large vessel occlusions in stroke cases performed solely by clinicians without AI support	Not mentioned. Use: computed angiography for stroke detection	People diagnosed with stroke (age 66 y)	71,840
3	Manual ASPECTS assessment for ischemic stroke diagnosis using computed tomography images sent via WhatsApp (Meta)	RAPID ASPECTS by iSchemaView. Use: automated cerebrovascular analysis in computed tomography images for stroke detection	Patients with ischemic stroke who received intravenous thrombolysis (average age 59 y)	122
4	The exclusive use of Lung-RADS classification by radiologists for risk stratification of lung nodules in baseline lung cancer screening	Not mentioned. Use: automated digital chest radiography for lung cancer detection	General population (63% men; average age 61 y; 93% White)	3197
5	The use of radiologists for reading chest x-rays in tuberculosis (TB) screening and triage	CAD4TB (Delft Imaging, Netherlands), Lunit INSIGHT CXR (Lunit, South Korea), qXR (Qure.ai, India), and InferRead DR Chest (Infervision, China). Use: automated digital chest radiography for TB detection	General population (no detailed information)	1.378 million
6	Standard of care tuberculosis (TB) and HIV screening without systematic digital chest x-ray and computer-aided detection (CAD)[e]	CAD4TBv5 Use: automated digital chest radiography for TB detection	Adults (>18 years) with cough symptoms (44% men; average age 33.8 y)	1462
7	Direct microbiological testing (smear microscopy or GeneXpert) for all individuals with presumptive TB, without previous triage	CAD4TBv6 and qXRv2 Use: automated digital chest radiography for TB detection	People with symptoms suggestive of TB (52% men; age 15 y and older, average 33 y)	N/A
8	Traditional radiologist interpretation of chest x-rays for pathology classification	Not mentioned. Use: automated digital chest radiography for pulmonary disease detection	General population (55% men, all ages, 36% [40-60] years, 34% [60-80] years)	129,819
9	Traditional radiologist interpretation of baseline low-dose CT[f] scans for lung cancer screening	Not mentioned. Use: automated digital chest radiography for lung cancer detection	General population (age 60 y)	N/A

[a]AI: artificial intelligence.

[b]Convolutional neural network (CNN), a type of machine learning model designed to analyze visual data, such as images, by automatically identifying patterns and features.

[c]N/A: not applicable.

[d]ASPECTS: Alberta stroke program early CT score.

[e]Computer-aided detection (CAD) refers to the use of computer algorithms to assist in the identification and diagnosis of medical conditions; CAD systems help health care professionals by highlighting potential areas of concern, improving diagnostic accuracy and efficiency.

[f]CT: computed tomography.

XSL•FO
RenderX

All studies explicitly stated the context of the study and the main question in the introduction. Moreover, they described a study design suitable for their objectives. However, the information on the populations studied was limited. Two studies [30,35] (25.0%) provided no further information, 3 [31,34,37] (37.5%) provided information on the age of participants, 3 [29,32,36] (37.5%) mentioned age and gender, and only 1 study [33] (12.5%) provided information on age, gender, and race or ethnicity. Regarding the use of AI, most studies referred to the technology's name (n=5, 62.5%). These technologies were mainly used to detect TB (n=3, 37.5%) [30-32], followed by applications in lung cancer [29,34] (n=2, 25.0%), stroke [35,36] (n=2, 25.0%), and 1 study [37] (12.5%) that focused on assisting in melanoma diagnosis.

Transferability Results

In terms of transferability, only 2 studies [35,37] (25%) were considered transferable to Brazil and other countries with a similar health context. Five studies [29-31,35,36] did not meet the discount rate criterion (62.5%), 1 study [36] did not address medical costs (12.5%), and the majority (75%) [30-32,34,36,37] did not meet the case mix criterion (Tables 4 and 5).

Equity Assessment

For the equity study, we used the EPHPP checklist and analyzed the study exhibited robust findings, with no sections identified as weak (Tables 4 and 5). The comprehensive selection of databases bolstered the representativeness of the study's population. The study design was also deemed appropriate for the methodology used. While the study could not analyze race or ethnicity and health insurance confounders for 2 of the 3 databases, it was able to assess gender and age in all cases. The data sources used were appropriately referenced by the authors and were considered reliable. The statistical methods used were deemed appropriate.

Economic Evaluation

Among the 8 economic evaluation studies identified, the majority (75%, N=6) focused on cost-effectiveness. A cost-effectiveness study evaluates the correlation between the financial resources invested and the benefits obtained, offering an essential analysis for enhancing the economic efficiency of an intervention, project, or program. However, 2 of the studies [29,30] (25.0%) were cost-minimization studies, which focus on identifying the most economical option among alternatives that yield similar outcomes, aiming to minimize expenses while delivering comparable results. Out of the 6 cost-effectiveness studies [31,32,34-37], 2 [34,35] (33.3%) examined the societal perspective, another 2 [31,32] (33.3%) considered the viewpoint of health care providers, and the remaining 2 [36,37] (33.3%) analyzed a combined perspective that encompasses both the health care system and society. Regarding the cost-minimization studies, one [29] adopted the health care system's perspective, while the other [30] took a combined approach involving both the health care system and society.

The results varied as the studies examined different scenarios with different methods. Nevertheless, the majority of the studies [29,31,32,34-37] (7 out of 8, 87.5%) showed positive results regarding the use of AI in the diagnosis of disease in dermatology, neurology, and pulmonology. The sole study presenting an unfavorable view pertained to the AI's efficacy in enhancing TB detection [30]. This singular negative outcome may be ascribed to several factors, one of which involves the economic implications of subsequent diagnostic tests required to validate the screening results, which were conducted in 61% of the study's participants. A summary of the methodological details of these economic evaluation studies is provided in Table 6.

XSL•FO
RenderX

Table . Details of the economic or equity evaluation of the studies.

ID	Model type	Perspective	Outcome measure	Direction of effect
1	Markov	Patient and health care system	The ICER[a] for dermatology and ophthalmology was –US $27,580 and –US $38,848 per QALY[b], respectively. The dominance in favor of AI[c] depended on small differences in the fee-for-service and assumed postdiagnosis treatment.	Favorable
2	Markov	Society	Over a lifetime horizon, AI resulted in cost savings of US $156 (0.23%) and an incremental gain of 0.01 QALYs per patient with suspected ischemic stroke. For each annual patient cohort, this translates to a total cost savings of US $11 million and a gain of 682 QALYs.	Favorable
3	Decision tree	Health care system and society	The incremental cost of ASPECTS[d] interpretation with WhatsApp support compared to automated ASPECTS interpretation was US $220.94. The estimated treatment decision cost was US $618.28 over a 90-day period[e].	Favorable
4	NA[f]	Health care system	Net cost savings from the AI-based management strategy of US $72 per screened patient in the first year after screening.	Favorable
5	NA	Health care system and society	The cost per screening for two CAD (computer-aided detection) software was lower than that of a radiologist (US $0.70-$0.93) for the high-yield scenarios studied. Projecting nationwide implementation, the additional cost of deploying CAD software was US $2.65 to US $19.3 million compared to US $23.97 million for human readers over a 4-year period.	Favorable
6	Not mentioned	Health care service provider	AI-assisted HIV and tuberculosis detection had an ICER of US $4620.47 per QALY, above Malawi's cost-effectiveness threshold, making it cost-ineffective compared to conventional screening within the 56-day trial period.	Unfavorable

XSL•FO
RenderX

ID	Model type	Perspective	Outcome measure	Direction of effect
7	Decision tree	Health care service provider	Screening strategies using AI-based chest radiographs compared to initial smear microscopy and GeneXpert projected cost savings of US $23,233 and US $34,346 per 1000 people, respectively, and prevented 3% - 4% and 4% of DALYs[g] over a 1-year time horizon.	Favorable
8	NA	NA	The algorithms showed underdiagnosis bias in three large public datasets based on gender, ethnicity or race, age, socioeconomic status, and intersectional subgroups.	Unfavorable
9	Markov	Society	At a willingness to pay of US $100,000 per QALY over a twenty-year analysis period, AI support for initial screening is cost-effective up to a cost of US $1240 per patient, and the ICER remains negative up to a threshold of US $68.	Favorable

[a]ICER: Incremental Cost-Effectiveness Ratio.

[b]QALY: Quality-Adjusted Life Years.

[c]AI: artificial intelligence.

[d]ASPECTS: Alberta stroke program early CT score.

[e]*: considering that £1 was equivalent to US $0.06349 on May 1, 2020 [36].

[f]NA: not applicable.

[g]DALY: disability-adjusted life years.

Equity Evaluation

Only a single study focusing on equity evaluation was identified. This study used extensive datasets containing images from over 129,000 patients. The demographics of these patients displayed an almost equal distribution across genders, with the majority aged between 40 years and 80 years. The authors developed various AI models for detecting anomalies in chest x-rays. The findings from the equity evaluation indicated signs of underdiagnosis in AI-based interpretations of these imaging examinations, especially within specific subpopulations. For a comprehensive summary of the methodological details of this equity evaluation study, refer to Table 6, under study ID: 8.

Discussion

Principal Findings

This study presents a novel systematic review that examines the economic and equity implications of AI in medical diagnostic imaging across dermatology, neurology, and pulmonology. It stands out from previous research [9] by incorporating equity considerations and expanding the scope of medical specialties beyond those traditionally studied. The findings of this systematic review suggest that AI integration could provide economic benefits to health care systems, as the majority of included studies indicate that the technology is cost-effective. However, only 1 study assessed AI from an equity perspective, highlighting the issue of underdiagnosis, particularly among subgroups defined by gender, ethnicity, and socioeconomic status. This underscores the need for further research to ensure AI models are both cost-effective and equitable in clinical practice.

An important observation from this review is the notable scarcity of studies dedicated to exploring the economic and equity impacts of AI in the diagnostic imaging of dermatological, neurological, and pulmonary diseases. The bulk of the research has centered on pulmonology, tackling conditions like TB and lung cancer, followed by neurology. Dermatology emerged as the least studied field despite the availability of numerous open datasets for training models on skin lesions. The generalizability of AI algorithms and their practical application in clinical settings remains an unresolved issue. There is a necessity for real-life clinical trials using AI algorithms to broaden their applicability in routine medical practice. Dermatology, in particular, poses a unique challenge due to the diversity of skin tones, which requires a varied dataset for effective algorithm training [38].

Moreover, this review has also underscored a remarkable deficiency in studies investigating the aspect of equity in AI

XSL•FO
RenderX

applications. This observation highlights the necessity for more inclusive research focused on equity-related issues. Furthermore, it is critical to acknowledge the potential influence of publication bias, where studies yielding positive outcomes are more likely to be published, thereby shaping the literature landscape [39].

A significant proportion of the studies included were conducted in high-income countries. Even studies involving low-income countries often collaborated with institutions from high-income countries. This trend raises important questions about the applicability of these findings in health care settings with limited resources. Some authors argue that the adoption of AI in low- and middle-income countries should follow a distinct path, rather than in high-income countries due to differences in staff, clinical experience, disease profiles, demographics, digital infrastructure, and available medical equipment. Health care facilities facing resource constraints require an integrated approach that encompasses clinical training, infrastructure enhancement, and a gradual introduction of AI technologies [40].

A key concern highlighted in this review is the insufficient detail provided about the AI tools used in the studies. Many articles did not disclose critical details such as the technology's specific name, architectural framework, training and validation datasets, manufacturer, availability, and licensing details. These omissions limit result reproducibility and hinder health care professionals' ability to implement AI technologies in clinical practice [41]. To address this gap, we emphasize the need for studies to follow established AI reporting guidelines, such as the MINIMAR (Minimum Information for Medical AI Reporting) framework, which advocates for transparency in AI model development and validation [42]. Furthermore, future studies should include standardized performance metrics, report biases in training data, and disclose the impact of image acquisition parameters, particularly in public health settings where data variability is significant. Establishing clear reporting standards and regulatory expectations will be crucial in ensuring AI models are both clinically applicable and equitably implemented.

On the other hand, the descriptions of the study populations often lacked essential demographic details, such as age, gender, and race, despite large sample sizes. This paucity of data on the study populations might impede the extrapolation of results across various demographic groups, especially when evaluating the impact of AI from an equity standpoint [41,43]. To address potential biases in AI, it is crucial for future studies to incorporate detailed demographic information and ensure training datasets are diverse across race, gender, socioeconomic status, and geographic regions. In addition, AI models should undergo fairness testing and be adjusted using bias-mitigation techniques, such as reweighting algorithms and adversarial debiasing. Policy makers should establish guidelines for transparent reporting of model performance across subpopulations and encourage collaborations between AI developers and public health institutions to ensure AI applications are equitable and contextually appropriate for diverse health care settings.

This review also identified a gap in the information regarding the timeframes during which the medical images were collected and analyzed. Understanding the timing of data collection and analysis is crucial for contextualizing the relevance of the findings. The dynamic nature of this field is reflected in the recent publication dates of the studies, indicating an evolving landscape where new technologies may emerge and existing ones may advance. Singh et al [44], 2021 reported a significant growth in the interest in medical applications of AI from 1998 to 2019, with a marked increase starting in 2015. Our systematic review, including articles regardless of publication date, found all relevant articles published post 2020, signifying a recent spike in scientific interest in evaluating AI technologies for potential health care integration.

The transferability of these technologies to Brazil's public health system presented significant challenges. Only 2 of the studies were considered relevant to the Brazilian context, a finding that resonates with similar observations in the global research community. This highlights the potential for applying these findings to other countries with comparable demographics and health care circumstances. Numerous studies from various countries have also reported challenges in adapting AI technologies to their unique health care systems. This situation underscores the necessity for crafting tailored strategies to facilitate the effective international transfer of medical technologies. To make AI interventions more effective and globally relevant, it is necessary to promote region-specific research that accounts for demographic, socioeconomic, cultural, and infrastructure variations [41].

In dermatology, real-world clinical trials are particularly critical due to the diversity of skin tones, environmental exposures, and disease presentations across populations. Data collection should be standardized and consider cultural and socioeconomic variables that may impact model performance. This can be achieved using mixed-methods research designs, including qualitative interviews and focus groups to identify context-specific factors, followed by quantitative observational studies (eg, cross-sectional and longitudinal cohort studies) to assess data quality and variable distribution. Pilot studies and feasibility trials are also valuable to test recruitment strategies, data collection workflows, and model deployment logistics in diverse settings. Data analysis should involve appropriate statistical methods, such as stratified analyses and interaction testing, to validate the model's effectiveness across different populations and clinical contexts. These methodological steps are essential for informing the design of robust, equitable clinical trials that evaluate AI-based interventions. In summary, future research should prioritize external validation of AI models on datasets that reflect social determinants of health to ensure equitable performance. In addition, partnerships between research institutions in low- and middle-income countries and international organizations could facilitate the adaptation and implementation of AI solutions, while also ensuring that developed models are more representative of the populations served. Incorporating standardized evaluation frameworks and transparency in AI reporting will further support the safe and effective deployment of dermatology-focused AI in diverse health care environments.

The only economic evaluation study reporting a disadvantageous outcome for AI use compared to standard care established a

XSL•FO
RenderX

cost-effectiveness threshold of up to US $1200 per QALY [31]. This threshold, variable across nations, is influenced by factors like per capita income, public preferences, resource availability, and health policy, yet has inherent limitations. For example, in Brazil, the threshold is set at US $7914 per QALY (considering that US $1 was equivalent to BRL R$5.05 on October 20, 2023), increasing to US $23,744 per QALY for severe and rare diseases [45]. In the United Kingdom, the range is between US $24,335 and US $36,502 per QALY gained (considering that US $1 was equivalent to GBP 0.82 on October 20, 2023) [46]. To better contextualize AI cost-effectiveness, it is essential to examine cost thresholds across different health care systems. Given the economic disparities between high-, middle-, and low-income countries, AI adoption must be evaluated within country-specific willingness-to-pay thresholds. In low- to middle-income countries, where health care budgets are more constrained, AI solutions must demonstrate cost savings that justify their implementation, particularly when compared to existing diagnostic methods. Policy makers should consider not only the direct costs of AI but also its long-term benefits, such as improved diagnostic accuracy, reduced hospitalizations, and enhanced workforce efficiency. Future research should explore region-specific economic models to determine realistic AI pricing structures that optimize health care accessibility and sustainability.

The studies in this review imply AI's potential to enhance cost-effectiveness, efficiency, and quality of life in health care, aligning with other research demonstrating AI as a cost-saving diagnostic tool. AI's acceleration of diagnosis, reduction of subjectivity, and facilitation of treatment allow health care systems to allocate resources more efficiently, yielding tangible benefits and cost reductions. Hence, AI applications are emerging as an innovative strategy potentially beneficial to patient care, with prospects for widespread adoption and long-term cost minimization [47].

However, in terms of equity, the sole study meeting the criteria indicated unfavorable outcomes, highlighting underdiagnosis in specific subgroups. This is significant as health equity is a core principle in health care, and AI should not perpetuate or intensify existing disparities. Other studies have also shown that, despite accuracy in some AI models, racial disparities can persist [18]. To ensure equity, it is recommended that AI models incorporate indicators of social determinants, such as environmental, occupational, and income exposures. Efforts should be made to integrate the full etiological "context" of a patient's health and optimize these models [43].

This study has limitations, including the potential bias of the chosen databases towards North American and European studies, possibly omitting relevant research from other regions, particularly Asia. Furthermore, the exclusion of gray literature, such as articles in open repositories like arXiv, might have constrained the scope of our review. Another significant challenge was the variability in study designs, data sources, and methodological approaches across the reviewed articles. This diversity made it difficult to synthesize the results comprehensively, rendering a meta-analysis unfeasible. Finally, the focus of this study on three specific medical fields—dermatology, neurology, and pulmonology—meant dealing with a limited number of available primary studies. This focus inherently excluded a vast array of studies in other medical areas, potentially overlooking relevant findings and insights that could be applicable to the broader discussion of AI in health care.

In summary, this systematic review emphasizes the imperative for a more robust and inclusive approach to using AI in medical imaging diagnostics, with particular attention to equity. Although the reviewed studies showcase AI's potential in enhancing medical diagnostics, it's crucial to ensure that such technologies do not deepen or perpetuate existing health disparities. Transparency in the depiction of AI tools and ensuring that study populations are representative are essential factors for consideration in future research endeavors. With AI rapidly integrating into health care, its comprehensive evaluation is necessary to guarantee that its benefits are accessible to all patients, irrespective of race, age, gender, or socioeconomic background. In addition, it is important to recognize that technologies developed within specific contexts, such as in high-income countries or for predominantly white populations, might not seamlessly transfer to other contexts or nations.

Conclusion

This review underscores the importance of transparency in the description of AI tools and representativeness of population subgroups to mitigate health disparities. As AI is rapidly being integrated into health care, detailed assessments are essential to ensure that benefits reach all patients, regardless of sociodemographic factors.

Acknowledgments

The authors would like to thank the people who contributed significantly to this project, including Pedro Vieira Santana Netto and the PROADI-SUS *Banco de Imagens* team; Leandro Seiti Anazawa, Tayna Gomes de Souza Bandeira, Renato Tanjoni, and their team; as well as the Edna Terezinha Rother's library team and the team from the Department of Science and Technology (SCTIE/DECIT) of the Brazilian Ministry of Health.

This project received funding from the Brazilian Ministry of Health, in collaboration with the Hospital Israelita Albert Einstein, under the *Programa de Apoio ao Desenvolvimento Institucional do Sistema Único de Saúde* (PROADI-SUS, NUP 25000.161106/2020 - 61).

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

GOS, LRC, and JPQP contributed to the conceptualization of the study. Data curation was performed by GOS, LRC, RMC, RML, and BCRS. Formal analysis was conducted by GOS, RMC, and LGNP. ETR provided the necessary resources. LRC supervised the project. The original draft of the manuscript was written by RMC and GOS. All authors contributed to reviewing and editing the manuscript.

Conflicts of Interest

The authors are involved in the PROADI-SUS *Banco de Imagens* project. This project represents a joint effort between the Brazilian Ministry of Health and Hospital Israelita Albert Einstein, aiming to improve public health in Brazil. The initiative's primary focus is on establishing a nationwide, cloud-based medical image repository and developing AI algorithms to support disease diagnosis.

Multimedia Appendix 1
Search strategy on economic evaluations or equity in the use of artificial intelligence tools for diagnostic support in imaging examinations in the dermatological area.
[DOCX File, 21 KB - i-jmr_v14i1e56240_app1.docx]

Multimedia Appendix 2
Search strategy on economic evaluations or equity in the use of artificial intelligence tools for diagnostic support in imaging examinations in the pulmonary area.
[DOCX File, 23 KB - i-jmr_v14i1e56240_app2.docx]

Multimedia Appendix 3
Search strategy on economic evaluations or equity in the use of artificial intelligence tools for diagnostic support in imaging examinations in the neurological area.
[DOCX File, 24 KB - i-jmr_v14i1e56240_app3.docx]

Multimedia Appendix 4
Search strategy on economic evaluations or equity in the use of artificial intelligence tools for diagnostic support in imaging examinations.
[DOCX File, 23 KB - i-jmr_v14i1e56240_app4.docx]

Multimedia Appendix 5
Summary of the articles selected for the study.
[DOCX File, 13 KB - i-jmr_v14i1e56240_app5.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File, 34 KB - i-jmr_v14i1e56240_app6.docx]

References

1. Kruk ME, Gage AD, Arsenault C, et al. High-quality health systems in the sustainable development goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):e1196-e1252. [doi: 10.1016/S2214-109X(18)30386-3] [Medline: 30196093]
2. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243. [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
3. Rong G, Mendez A, Bou Assi E, Zhao B, Sawan M. Artificial intelligence in healthcare: review and prediction case studies. Engineering (Beijing) 2020 Mar;6(3):291-301. [doi: 10.1016/j.eng.2019.08.015] [Medline: 32837747]
4. McCarthy J. From here to human-level AI. Artif Intell 2007 Dec;171(18):1174-1182. [doi: 10.1016/j.artint.2007.10.009]
5. Doupe P, Faghmous J, Basu S. Machine learning for health services researchers. Value Health 2019 Jul;22(7):808-815. [doi: 10.1016/j.jval.2019.02.012] [Medline: 31277828]
6. Mitchell T, Buchanan B, DeJong G, Dietterich T, Rosenbloom P, Waibel A. Machine Learning. Annu Rev Comput Sci 1990 Jun;4(1):417-433. [doi: 10.1146/annurev.cs.04.060190.002221]

XSL•FO
RenderX

7. Nichols JA, Herbert Chan HW, Baker MAB. Machine learning: applications of artificial intelligence to imaging and diagnosis. Biophys Rev 2019 Feb;11(1):111-118. [doi: 10.1007/s12551-018-0449-9] [Medline: 30182201]
8. Generating Evidence for Artificial Intelligence-Based Medical Devices: A Framework for Training, Validation and Evaluation: Organization WHO; 2021.
9. Voets MM, Veltman J, Slump CH, Siesling S, Koffijberg H. Systematic review of health economic evaluations focused on artificial intelligence in healthcare: the tortoise and the cheetah. Value Health 2022 Mar;25(3):340-349. [doi: 10.1016/j.jval.2021.11.1362] [Medline: 35227444]
10. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med 2019 Oct 29;17(1):195. [doi: 10.1186/s12916-019-1426-2] [Medline: 31665002]
11. Siontis GCM, Sweda R, Noseworthy PA, Friedman PA, Siontis KC, Patel CJ. Development and validation pathways of artificial intelligence tools evaluated in randomised clinical trials. BMJ Health Care Inform 2021 Dec;28(1):e100466. [doi: 10.1136/bmjhci-2021-100466] [Medline: 34969668]
12. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. JAMA 2020 Sep 22;324(12):1212-1213. [doi: 10.1001/jama.2020.12067] [Medline: 32960230]
13. Corti C, Cobanaj M, Dee EC, et al. Artificial intelligence in cancer research and precision medicine: Applications, limitations and priorities to drive transformation in the delivery of equitable and unbiased care. Cancer Treat Rev 2023 Jan;112:102498. [doi: 10.1016/j.ctrv.2022.102498] [Medline: 36527795]
14. Mittermaier M, Raza MM, Kvedar JC. Bias in AI-based models for medical applications: challenges and mitigation strategies. NPJ Digit Med 2023;6(1):113. [doi: 10.1038/s41746-023-00858-z]
15. Lee H, Kim D, Lee S, Fawcett J. The concepts of health inequality, disparities and equity in the era of population health. Appl Nurs Res 2020 Dec;56:151367. [doi: 10.1016/j.apnr.2020.151367] [Medline: 33280788]
16. Cirillo D, Catuara-Solarz S, Morey C, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. NPJ Digit Med 2020;3(1):81. [doi: 10.1038/s41746-020-0288-5] [Medline: 32529043]
17. Clark CR, Wilkins CH, Rodriguez JA, et al. Health care equity in the use of advanced analytics and artificial intelligence technologies in primary care. J Gen Intern Med 2021 Oct;36(10):3188-3193. [doi: 10.1007/s11606-021-06846-x] [Medline: 34027610]
18. 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]
19. Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: implications for health systems. J Glob Health 2019 Dec;9(2):010318. [doi: 10.7189/jogh.09.020318] [Medline: 31788229]
20. Menicucci TMG. O Sistema Único de Saúde, 20 anos: balanço e perspectivas. Cad Saúde Pública 2009 Jul;25(7):1620-1625. [doi: 10.1590/S0102-311X2009000700021]
21. Celuppi IC, Lima GDS, Rossi E, Wazlawick RS, Dalmarco EM. An analysis of the development of digital health technologies to fight COVID-19 in Brazil and the world. Cad Saude Publica 2021;37(3):e00243220. [doi: 10.1590/0102-311X00243220] [Medline: 33729283]
22. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Int J Surg 2021 Apr;88:105906. [doi: 10.1016/j.ijsu.2021.105906] [Medline: 33789826]
23. Santana GO, Couto RDM, Loureiro RM, et al. Economic evaluations and equity in the use of artificial intelligence in imaging exams for medical diagnosis in people with skin, neurological, and pulmonary diseases: protocol for a systematic review. JMIR Res Protoc 2023 Dec 28;12:e48544. [doi: 10.2196/48544] [Medline: 38153775]
24. Welch V, Petticrew M, Tugwell P, et al. PRISMA-Equity 2012 extension: reporting guidelines for systematic reviews with a focus on health equity. PLoS Med 2012;9(10):e1001333. [doi: 10.1371/journal.pmed.1001333] [Medline: 23222917]
25. World bank group country classifications by income level for FY24 (july 1, 2023- june 30, 2024). World Bank. 2023. URL: https://blogs.worldbank.org/opendata/new-world-bank-group-country-classifications-income-level-fy24 [accessed 2025-07-03]
26. Wijnen B, Van Mastrigt G, Redekop WK, Majoie H, De Kinderen R, Evers S. How to prepare a systematic review of economic evaluations for informing evidence-based healthcare decisions: data extraction, risk of bias, and transferability (part 3/3). Expert Rev Pharmacoecon Outcomes Res 2016 Dec;16(6):723-732. [doi: 10.1080/14737167.2016.1246961] [Medline: 27762640]
27. 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]
28. Welte R, Feenstra T, Jager H, Leidl R. A decision chart for assessing and improving the transferability of economic evaluation results between countries. Pharmacoeconomics 2004;22(13):857-876. [doi: 10.2165/00019053-200422130-00004] [Medline: 15329031]
29. Adams SJ, Mondal P, Penz E, Tyan CC, Lim H, Babyn P. Development and cost analysis of a lung nodule management strategy combining artificial intelligence and Lung-RADS for baseline lung cancer screening. J Am Coll Radiol 2021 May;18(5):741-751. [doi: 10.1016/j.jacr.2020.11.014] [Medline: 33482120]

XSL•FO
RenderX

30. Bashir S, Kik SV, Ruhwald M, et al. Economic analysis of different throughput scenarios and implementation strategies of computer-aided detection software as a screening and triage test for pulmonary TB. PLoS ONE 2022;17(12):e0277393. [doi: 10.1371/journal.pone.0277393] [Medline: 36584194]
31. MacPherson P, Webb EL, Kamchedzera W, et al. Computer-aided X-ray screening for tuberculosis and HIV testing among adults with cough in Malawi (the PROSPECT study): a randomised trial and cost-effectiveness analysis. PLoS Med 2021 Sep;18(9):e1003752. [doi: 10.1371/journal.pmed.1003752] [Medline: 34499665]
32. Nsengiyumva NP, Hussain H, Oxlade O, et al. Triage of persons with tuberculosis symptoms using artificial intelligence-based chest radiograph interpretation: a cost-effectiveness analysis. Open Forum Infect Dis 2021 Dec;8(12):ofab567. [doi: 10.1093/ofid/ofab567] [Medline: 34917694]
33. Seyyed-Kalantari L, Zhang H, McDermott MBA, Chen IY, Ghassemi M. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. Nat Med 2021 Dec;27(12):2176-2182. [doi: 10.1038/s41591-021-01595-0] [Medline: 34893776]
34. Ziegelmayer S, Graf M, Makowski M, Gawlitza J, Gassert F. Cost-effectiveness of artificial intelligence support in computed tomography-based lung cancer screening. Cancers (Basel) 2022 Mar 29;14(7):1729. [doi: 10.3390/cancers14071729] [Medline: 35406501]
35. van Leeuwen KG, Meijer FJA, Schalekamp S, et al. Cost-effectiveness of artificial intelligence aided vessel occlusion detection in acute stroke: an early health technology assessment. Insights Imaging 2021 Sep 25;12(1):133. [doi: 10.1186/s13244-021-01077-4] [Medline: 34564764]
36. Mansour OY, Ramadan I, Abdo A, et al. Deciding thrombolysis in AIS based on automated versus on WhatsApp interpreted ASPECTS, a reliability and cost-effectiveness analysis in developing system of care. Front Neurol 2020;11:333. [doi: 10.3389/fneur.2020.00333] [Medline: 32508730]
37. Gomez Rossi J, Rojas-Perilla N, Krois J, Schwendicke F. Cost-effectiveness of artificial intelligence as a decision-support system applied to the detection and grading of melanoma, dental caries, and diabetic retinopathy. JAMA Netw Open 2022 Mar 1;5(3):e220269. [doi: 10.1001/jamanetworkopen.2022.0269] [Medline: 35289862]
38. Liopyris K, Gregoriou S, Dias J, Stratigos AJ. Artificial intelligence in dermatology: challenges and perspectives. Dermatol Ther (Heidelb) 2022 Dec;12(12):2637-2651. [doi: 10.1007/s13555-022-00833-8] [Medline: 36306100]
39. Siddaway AP, Wood AM, Hedges LV. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Annu Rev Psychol 2019 Jan 4;70:747-770. [doi: 10.1146/annurev-psych-010418-102803] [Medline: 30089228]
40. Mollura DJ, Culp MP, Pollack E, et al. Artificial intelligence in low- and middle-income countries: innovating global health radiology. Radiology 2020 Dec;297(3):513-520. [doi: 10.1148/radiol.2020201434] [Medline: 33021895]
41. McCoy LG, Banja JD, Ghassemi M, Celi LA. Ensuring machine learning for healthcare works for all. BMJ Health Care Inform 2020 Nov;27(3):e100237. [doi: 10.1136/bmjhci-2020-100237] [Medline: 33234535]
42. Hernandez-Boussard T, Bozkurt S, Ioannidis JPA, Shah NH. MINIMAR (MINimum Information for Medical AI Reporting): Developing reporting standards for artificial intelligence in health care. J Am Med Inform Assoc 2020 Dec 9;27(12):2011-2015. [doi: 10.1093/jamia/ocaa088] [Medline: 32594179]
43. Dankwa-Mullan I, Weeraratne D. Artificial intelligence and machine learning technologies in cancer care: addressing disparities, bias, and data diversity. Cancer Discov 2022 Jun 2;12(6):1423-1427. [doi: 10.1158/2159-8290.CD-22-0373] [Medline: 35652218]
44. Singh S, Tiwari R, Tiwari G. Importance of artificial intelligence in the medical device and health care sector. Pharma Times 2021;53(11):21.
45. Brasil. Ministério da Saúde. Secretaria de Ciência, Tecnologia, Inovação e Insumos Estratégicos em Saúde. Departamento de Gestão e Incorporação de Tecnologias em Saúde. Coordenação Geral de Avaliação de Tecnologias em Saúde, Secretaria de Ciência, Tecnologia, Inovação e Insumos Estratégicos em Saúde, Departamento de Gestão e Incorporação de Tecnologias em Saúde, Coordenação Geral de Avaliação de Tecnologias em Saúde. O uso de limiares de custo-efetividade nas decisões em saúde: recomendações da comissão nacional de incorporação de tecnologias no SUS [article in Portuguese]. : Ministério da Saúde; 2022.
46. Nacional institute for health and care excellence. : NICE health technology evaluations: the manual; 2022.
47. Khanna NN, Maindarkar MA, Viswanathan V, et al. Economics of artificial intelligence in healthcare: diagnosis vs. treatment. Healthcare (Basel) 2022 Dec 9;10(12):2493. [doi: 10.3390/healthcare10122493] [Medline: 36554017]

Abbreviations

AI: artificial intelligence
CHEC: Consensus on Health Economic Criteria
EPHPP: Effective Public Health Practice Project
ICER: incremental cost-effectiveness ratio
MINIMAR: Minimum Information for Medical AI Reporting
PICO: Population, Intervention, Comparison, and Outcome

XSL•FO
RenderX

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-E: Preferred Reporting Items for Systematic Reviews and Meta-Analyses - Equity Extension
PROADI-SUS: Programa de Apoio ao Desenvolvimento Institucional do Sistema Único de Saúde
PROSPERO : International Prospective Register of Systematic Reviews
QALY: quality-adjusted life year
TB: tuberculosis

Edited by TDA Cardoso, X Ma; submitted 10.01.24; peer-reviewed by E Bai, MJ Gonzalez, O El-Gayar, ST Arasteh; revised version received 13.05.25; accepted 14.05.25; published 13.08.25.

Please cite as:
Santana GO, Couto RDM, Loureiro RM, Furriel BCRS, de Paula LGN, Rother ET, de Paiva JPQ, Correia LR
Economic Evaluations and Equity in the Use of Artificial Intelligence in Imaging Examinations for Medical Diagnosis in People With Dermatological, Neurological, and Pulmonary Diseases: Systematic Review
Interact J Med Res 2025;14:e56240
URL: https://www.i-jmr.org/2025/1/e56240
doi:10.2196/56240

© Giulia Osório Santana, Rodrigo de Macedo Couto, Rafael Maffei Loureiro, Brunna Carolinne Rocha Silva Furriel, Luis Gustavo Nascimento de Paula, Edna Terezinha Rother, Joselisa Péres Queiroz de Paiva, Lucas Reis Correia. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 13.8.2025. 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 for Improving Body Dissatisfaction in Children and Emerging Adults: Systematic Review and Meta-Analysis

Li Liu[1,2], BSc; Jianning Yang[3], MEng; Fengmei Tan[2], MSN; Xia Yang[2], BSN; Huan Luo[4], BSN; Yanhua Chen[4], PhD; Xiaolei Zhao[1], PhD

[1]School of Nursing, Southwest Medical University, 1 Xianglin Road, Longmatan District, Luzhou, Sichuan, China

[2]Department of Hematopathology, The Affiliated Hospital, Southwest Medical University, Luzhou, Sichuan, China

[3]Department of Nephrology, The Affiliated Hospital, Southwest Medical University, Luzhou, Sichuan, China

[4]Department of Nursing, The Affiliated Hospital, Southwest Medical University, Luzhou, Sichuan, China

Corresponding Author:
Xiaolei Zhao, PhD
School of Nursing, Southwest Medical University, 1 Xianglin Road, Longmatan District, Luzhou, Sichuan, China

Abstract

Background: Body dissatisfaction is a condition where individuals are dissatisfied with their physical appearance. It has become a global issue, especially among children and emerging adults. A growing number of digital interventions have been developed to address body dissatisfaction in children and emerging adults; however, controversies remain regarding their efficacy, underscoring the need for a comprehensive synthesis of current evidence.

Objective: This systematic review aimed to explore the effectiveness of digital interventions in improving body image–related outcomes among children and emerging adults.

Methods: From inception to April 24, 2024, a literature search was performed across 7 databases—PubMed, Web of Science, MEDLINE, EBSCO (Elton B Stephens Company), Cochrane Library, CNKI (China National Knowledge Infrastructure), and WANFANG—to identify randomized controlled trials (RCTs) with a predefined set of inclusion criteria. This systematic review was reported in line with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines. Study selection, data extraction, and risk of bias assessment using the Cochrane Risk-of-Bias Tool 2.0 were conducted independently by 2 researchers. Standardized mean differences (SMDs) and 95% CIs from the included RCTs were calculated for the meta-analysis. Heterogeneity was assessed with I^2 values. A fixed-effects model was used when $I^2 \leq 50\%$, and a random-effects model was selected when $I^2 > 50\%$.

Results: Twenty RCTs with 5251 participants (2610 in intervention groups and 2641 in control groups) met the inclusion criteria. Digital interventions included web pages, mobile apps, computer-based videos, computer-based sessions, internet-based sessions, internet games, chatbots, podcasts, and social media. Our results indicate that digital interventions could significantly improve body dissatisfaction (SMD=0.38, 95% CI −0.63 to −0.13; I^2=55%; P=.003), physical appearance comparison (SMD=−0.24, 95% CI −0.45 to −0.03; I^2=0%; P=.003), thin-ideal internalization (SMD=−0.28, 95% CI −0.36 to −0.2; I^2=41%; P<.001), self-esteem (SMD=0.14, 95% CI 0.07-0.21; I^2=21%; P<.001), self-compassion (SMD=0.55, 95% CI 0.33-0.78; I^2=35%; P<.001), and depression (SMD=−0.59, 95% CI −0.97 to −0.21; I^2=0%; P=.002), with small to medium effect sizes.

Conclusions: While digital interventions improved body dissatisfaction among children and emerging adults, additional well-designed, rigorous, and large-scale RCTs are needed to decisively provide estimates of the effectiveness of digital interventions on body dissatisfaction.

Trial Registration: PROSPERO CRD42024567594; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024567594

(Interact J Med Res 2025;14:e72231) doi:10.2196/72231

KEYWORDS

body dissatisfaction; digital intervention; children; emerging adults; systematic review; meta-analysis; PRISMA

XSL•FO
RenderX

Introduction

Prevalence and Impacts of Body Dissatisfaction

Body image encompasses an individual's cognition, emotional attitudes, and behavioral regulation regarding their body, and is a multidimensional concept [1]. Dissatisfaction with one's body's appearance, including its shape, weight, and other features, is known as body dissatisfaction [2]. Body dissatisfaction has a pervasive hold on children and emerging adults, and the phenomenon has been described as "normative discontent" [3]. Studies showed the prevalence of body dissatisfaction ranged from 42.2% to 80.9% among children and emerging adults across the world [4-7]. Females, those with higher BMI, and those from minority groups were reported to experience greater body dissatisfaction [8]. Body dissatisfaction can lead to a range of physiological issues, such as low self-esteem [9], anxiety [10], and depression [11], as well as behavioral burdens, such as eating disorders [12], extreme weight loss behaviors [13], alcohol, drug abuse [14], and excessive pursuit of cosmetic procedures [15]. The direct economic costs of body dissatisfaction were estimated to be US $84 billion in the United States [16].

Existing Face-to-Face Interventions and Limitations

Existing face-to-face interventions, such as cognitive behavioral therapy (CBT) and behavior interventions and supports (BIS), demonstrated small to moderate efficacy in reducing body dissatisfaction. For instance, CBT helps reframe maladaptive thoughts about body image [17], while dissonance-based interventions like the EVERYbody Project reduced thin-ideal internalization and eating disorder symptoms in college students [18]. A systematic review revealed that behavior interventions and supports improved disordered eating, body dissatisfaction, and extreme weight control behaviors in girls, but not boys [19]. However, the face-to-face interventions are sometimes constrained by accessibility, the costly nature of the format, a global shortage of mental health professionals, or stigma [20].

Efficacy of Digital Interventions

Digital technologies offered cost-effective alternatives in body dissatisfaction interventions [21]. Although studies identified that social media engagement was associated with higher body dissatisfaction and restricting food [22], digital interventions could offer especially critical support in adolescents [23]. A systematic review assessed the effectiveness of universal body image interventions delivered through a digital platform among young women. Most articles indicated that these interventions were effective in improving at least one body image outcome [24]. However, findings in other studies remained inconsistent. A study involving 127 girls aged 10-13 years, using videos developed by the Dove Self-Esteem Project, revealed that after intervention, there was no significant difference between the intervention group and the control group on body satisfaction [25]. This highlights the need for a comprehensive synthesis of current evidence.

Gaps in Current Evidence

There was no systematic review to explore the effectiveness of digital interventions on body dissatisfaction in children and emerging adults through quantitative synthesis. In this regard, this study aimed to fill this identified gap in the literature through a meta-analysis and to synthesize the effectiveness of various digital interventions on children and emerging adults with body dissatisfaction. By providing meta-analyses of currently available randomized controlled trials in children and emerging adults, this paper aimed to consolidate evidence on the use of digital interventions to treat body dissatisfaction.

Methods

Search Strategy

This systematic review was reported in line with PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis) guidelines [26] (the PRISMA checklist is provided in Checklist 1). PubMed, Web of Science, MEDLINE, EBSCO (Elton B Stephens Company), Cochrane Library, the Chinese databases CNKI (China National Knowledge Infrastructure) and WANFANG were searched from inception to April 24, 2024. This process was carried out independently by 2 researchers (LL & JNY). The PICOS (population, intervention, comparison, outcome, and study design) formula was used:

- Population (#1): youth OR young* OR child* OR adolescent OR teen* OR juvenile OR junior OR girl OR boy OR adult or students;
- Intervention (#2): remote OR website OR digital OR online OR network OR phone OR internet OR eHealth OR mHealth OR app OR multimedia OR social media OR zoom OR Facebook OR Instagram OR telephone OR (virtual reality);
- Comparison (#3): "usual care";
- Outcomes (#4): (body image) OR (body shape) OR (body dissatisfaction) OR (body weight) OR (physical appearance);
- Study design (#5): randomized controlled trials.

The final search strategy is #1 AND #2 AND #3, with #4 (Type of Study: randomized controlled trial) applied as a filter. Reference lists of publications were also searched for potentially relevant studies (see Multimedia Appendix 1).

Inclusion and Exclusion Criteria

Design

Any randomized controlled trial (RCT) in English or Chinese that explored the effects of digital interventions on body dissatisfaction among children or emerging adults was included, including randomized waitlist-controlled trials and crossover RCTs.

Participants

Children (<18 years [27]) or emerging adults (18 - 25 years [28]) with body dissatisfaction were included. Some special groups of children or emerging adults were excluded, including but not limited to those who were pregnant, postpartum women, new mothers, children, or emerging adults with amputation, patients with cancer, models, or cosmetic surgery sequelae. Only children's data were extracted if parents and children had been involved in the RCTs as participants.

XSL•FO
RenderX

Intervention

Intervention methods were based on Internet or smartphone technologies, including apps, web pages, virtual reality, telemedicine, Zoom (Zoom Video Communications), Facebook (Meta Platforms), Instagram (Meta Platforms), and other social media.

Outcomes

Outcomes of interest include at least one of the following measures, primary outcomes: body image satisfaction or dissatisfaction, such as body dissatisfaction, body appreciation, shape and weight concerns, and physical appearance comparisons; secondary outcomes: internalized outcomes, such as self-esteem, self-compassion, thin-ideal internalization, and self-objectification; negative affect, such as depression and anxiety symptoms, and negative affect; and eating behaviors, such as eating disorder and eating restraint. These secondary outcomes were selected based on the Tripartite Influence Model [29], which presents that family, peers, and media influence an individual's body dissatisfaction via appearance-related social comparisons and thin-ideal internalization. Thin-ideal internalization serves as a predictor of body dissatisfaction [30], while studies showed that body dissatisfaction acted as a risk factor for negative affect [31] and eating disorder pathology [32].

Study Selection

The reference manager EndNote 21 (Clarivate) was used to manage studies and remove duplicates. The remaining records were screened by 2 independent reviewers (LL and JNY) based on the titles and the abstracts. The full text of studies that either reviewer identified as potentially eligible continued to be read and screened by the 2 independent reviewers (LL and JNY) based on the inclusion and exclusion criteria. Any discrepancies were resolved with the third reviewer (XLZ) until consensus was reached.

Quality Assessment

The risk of bias in eligible RCTs was independently assessed by the 2 reviewers (LL and JNY) according to 7 domains of the Cochrane risk-of-bias tool for randomized trials [33]: (1) random sequence generation, (2) allocation concealment, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data, (6) selective reporting, and (7) other potential sources of bias. For each domain, the risk of bias was classified as low, high, or unclear. Discrepancies were adjudicated by the third reviewer (XLZ) until a consensus was achieved. Studies were considered low risk of bias if all 7 domains were assessed as low risk, or only one domain was assessed as high risk or unclear. If 2 domains were assessed as high risk or unclear, the studies were determined to have some concerns. If more than 2 domains were assessed as high risk or unclear, the study was rated a high risk of bias [34].

Data Extraction

A standardized data extraction form was used to extract data from each study, which included the following details: (1) first author, the year of publication, and country; (2) inclusion and exclusion criteria; (3) sample size of the experimental and control groups; (4) intervention characteristics (contents of intervention and comparison, follow-up); and (5) outcomes and main results. Data were extracted by LL and JNY and verified by XLZ. Only the data in the first period were extracted in the randomized waitlist-controlled trials. Only the data in the digital intervention groups and the control groups were extracted when participants were divided into more than 2 groups.

Data Analysis

Review Manager v5.4.1 (Cochrane Collaboration) was used for data synthesis. The findings that could not be synthesized were narratively described based on the outcomes of interest. For the continuous variables, standardized mean difference (SMD) and 95% CIs were calculated through random or fixed-effects models when the studies assessed the same outcome. SMD values of 0.2 - 0.5 represented a small effect size, 0.5 - 0.8 was considered medium, and values greater than 0.8 were interpreted as large [35]. The effect size of individual studies was mainly combined using the random effects model because of the different scales involved in assessing the same outcomes. The I^2 statistic was used to assess the heterogeneity across studies. I^2 values of 25%, 50%, and 75% were considered as low, moderate, and high heterogeneity [36]. When $I^2 \leq 50\%$, a fixed-effects model was selected; otherwise, a random-effects model was used [37]. Strategies for addressing heterogeneity included choosing a random or fixed effect model, excluding studies, or conducting subgroup analysis. The test level was $\alpha=.05$, and $P<.05$ was considered to indicate statistical significance.

Ethical Considerations

This systematic review is based on the synthesis of previously published studies and does not involve the collection of primary data directly from human participants. As such, formal ethical approval, including an institutional review board approval number, informed consent, and compensation, was not applicable. However, we have adhered to the ethical principles of research, including the accurate and transparent reporting of study findings. We have also ensured that all included studies were conducted in accordance with relevant ethical guidelines and regulations.

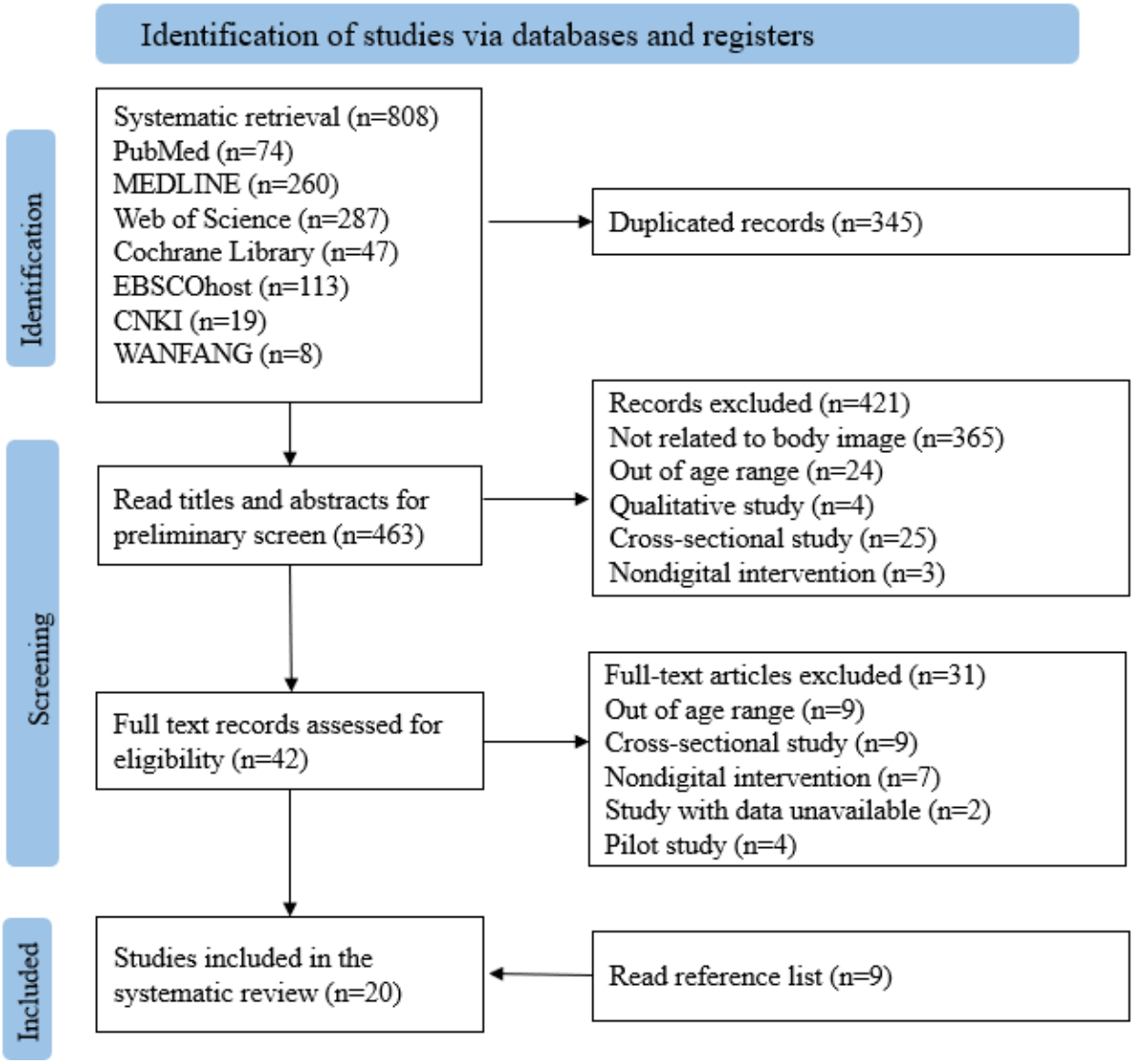
Results

Search Results

A total of 808 records were identified from the initial literature search. After removing the duplicates, titles and abstracts were screened, and 42 studies were further reviewed in full text. Finally, 20 studies [25,38-56] met the inclusion criteria, including 9 [41-43,49,50,53-56] studies identified from reference lists. The study selection flow chart is shown in Figure 1.

XSL•FO
RenderX

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram.



Study Characteristics

All the studies included were RCTs, including randomized waitlist-controlled trials (n=7). Only the data in the first phase were extracted in the randomized crossover and waitlist trials. Of the included studies, 20 RCTs with 5251 participants were included, including 2610 participants in the intervention groups and 2641 participants in the control groups. The studies were published between 2000 and 2023 and were conducted in the United States (n=10), Australia (n=3), the United Kingdom (n=3), China (n=1), Indonesia (n=1), Italy (n=1), and Brazil (n=1). Five studies [38-42] had mixed-gender samples, 15 [25,42-56] with only females. Sample sizes of the included studies ranged from 53 to 2000. All studies reported postintervention effects, whereas short-term and long-term follow-up data were reported in 11 [38,42,43,45,49,51-56] and 3 [39-41] studies, respectively. Participants and study characteristics are summarized in Table 1.

The digital interventions mainly included web page (n=6), computer-based sessions (n=1), internet-based sessions (n=2), computer-based videos (n=2), social media–based video (n=1), social media (n=3), such as Facebook and podcasts, mobile apps (n=2), internet games (n=1), animated films (n=1), and chatbot (n=1). Among the included studies, the interventions varied in length from 1 minute to 2 hours per session and in duration, ranging from a single day up to 8 weeks.

XSL•FO
RenderX

Table . Characteristics of the 20 included randomized controlled trials.

Author (year, country)	Participants	Sample size	Intervention	Comparison	Follow-up	Outcomes
Winzelberg et al,2000 (United States) [43]	Female university students	T:[a] n=31, C:[b] n=29	Internet-based sessions	Waitlist control group	3 months	BSQ[c], EDI-DT[d], EDEQ[e], WCSC[f]
Bruning et al, 2004 (United States) [44]	14 - to 16-year-old girls and their parents	T: n=102, C: n=51	Web page	Waitlist control group	6 months	WDEQ, WSCS[f], EDI-DT
Low et al, 2006 (United States) [45]	First- and second-year female undergraduates	T: n=14, C: n=14	Computer-Based sessions	No intervention	8 months	EDI[g], SFRS[h], EDI-DT, SATA[i]
Heinicke et al, 2007 (Australia) [46]	12 - to 18-year-old females	T: n=36, C: n=37	Internet-based sessions	Waitlist control group	2 months, 6 months	BSQ-SF[j], BCS[k], DEBQ[l], SATA, BDI[m]
Cousineau et al, 2010 (United States) [38]	6th-grade students	T: n=92, C: n=98	Web page	Science-based websites	3 months	BESAA[n], SPPA[o]
Halliwell et al, 2011 (United Kingdom) [25]	10 - to 13-year-old females	T: n=37, C: n=29	Computer-based videos	Newspaper or magazine	None	EDI-BD[p], BISS[q], BES[r], DEBQ[l], EDI-DT
Franko et al, 2012 (United States) [47]	College women	T: n=32, C: n=32	Web page	Websites without information about eating	3 months	BSQ, SATAQ-3[s]
Stice et al, 2012 (United States) [48]	Female college students	T: n=19, C: n=20	Web page	Educational brochure group: brochure	None	BPS-SD[t], IBSS[u]-R, DRES[v], BDI, EDDI[w]
Serdar et al, 2014 (United States) [49]	18 - 25 years old females	T: n=112, C: n=114	Web page	No intervention	8 - 9 weeks	BES, EDDS[x], IBSS-R[y]
Zhong et al, 2016 (China) [50]	Female college students	T: n=33, C: n=31	Web page	No intervention	None	BPS-SD, IBSS-R
Toole et al, 2016 (United States) [51]	18 - to 21-year-old females	T: n=40, C: n=40	Social media	Waitlist control group	None	BSQ, SCS[z], BAS[aa], RSES[ab], CSWS-AS[ac]
Slater et al,2017 (United Kingdom) [52]	8 - to 9-year-old females	T: n=40, C: n=40	Internet games	Internet games without human figures	None	CFRS[ad], MTST[ae]
Rodgers et al, 2018 (United States) [39]	14 - to 18-year-old adolescents	T: n=129, C: n=123	Mobile app	No intervention	12 weeks	PNSC[af], SCS[ag], BESAA[ah], PAC[ai]
Matheson et al, 2020 (United States) [40]	7 - to 14-year-old children	T: n=442, C: n=446	Animated films	Animation without body image	None	VAS[aj], CFRS
Seekis et al, 2020 (Australia) [53]	17 - to 21-year-old females	T: n=42, C: n=34	Social media	Waitlist control group	1 month, 3 months	EDI-BD, EDI-DT, SAAS[ak], UPACS[al], BAS-2[am], SCS-SF
Atkinson et al, 2021 (United Kingdom) [54]	Female undergraduates	T: n=67, C: n=65	computer-based videos	Documentary	1 week	VAS, BAS-2, BIAAQ[an], SATAS[ao], EDEQ, WSC[ap]
Cerea et al, 2021 (Italy) [41]	20 - to 25-year-old females	T: n=25, C: n=25	Mobile app	Waitlist control group	16 days	QDC[aq], DASS-21[ar], EDI-3[as]

XSL•FO
RenderX

Author (year, country)	Participants	Sample size	Intervention	Comparison	Follow-up	Outcomes
Garbett et al, 2023 (Indonesia) [55]	15 - to 19-year-old females	T: n=924, C: n=923	Social media–based (Facebook and Instagram) videos	Waitlist control group	1 month	BESAA, SATAQ[at], PNASC
Matheson et al, 2023 (Brazil) [42]	13 - to 18-year-old Brazilian residents	T: n=355, C: n=443	Chatbot	Standard care	1 week, 1 month	BESAA, PNASC, BIS-ES[au]

XSL•FO
RenderX

Author (year, country)	Participants	Sample size	Intervention	Comparison	Follow-up	Outcomes
Fardouly et al, 2023 (Australia) [56]	18 - to 25-year-old females	T: n=38, C: n=47	Social media	Use Facebook as usual	4 weeks	EDI-BD, PNASC, BAS, PACS

[a]T: intervention group.
[b]C: control group.
[c]BSQ: Body Shape Questionnaire.
[d]EDI-DT: Eating Disorder Inventory-Drive for Thinness.
[e]EDEQ: Eating Disorder Examination Questionnaire.
[f]WCSC: Weight Concerns and Shape Concerns scale.
[g]EDI: Eating Disorders Inventory.
[h]SFRS: The Stunkard Figure Rating Scale.
[i]SATA: Sociocultural Attitudes Toward Appearance.
[j]BSQ-SF: Body Shape Questionnaire-Short form.
[k]BCS: Body Comparison Scale.
[l]DEBQ: Dutch Eating Behavior Questionnaire.
[m]BDI: Beck Depression Inventory.
[n]BESAA: Body Esteem Scale for Adolescents and Adults.
[o]SPPA: Self-Perception Profile for Adolescents
[p]EDI-BD: Body Dissatisfaction Subscale of the Eating Disorder Inventory.
[q]BISS: Body Image State Scale.
[r]BES: Body Esteem Scale.
[s]SATAQ-3: Sociocultural Attitudes Towards Appearance Questionnaire-3.
[t]BPS-SD: Satisfaction and Dissatisfaction with Body Parts Scale.
[u]IBSS: Ideal-Body Stereotype Scale-Revised.
[v]DRES: Dutch Restrained Eating Scale
[w]EDDI: Eating Disorder Diagnostic Interview.
[x]EDDS: Eating Disorder Diagnostic Scale.
[y]IBSS-R: Ideal-Body Stereotype Scale-Revised.
[z]SCS: Self-Compassion Scale.
[aa]BAS: Body Appreciation Scale.
[ab]RSES: Rosenberg Self-Esteem Scale.
[ac]CSWS-AS: Contingencies of Self-Worth Scale-Appearance Subscale.
[ad]CFRS: Child Figure Rating Scale.
[ae]MTST: Modified Twenty Statements Test.
[af]PNASC: Positive and Negative Affect Schedule for Children.
[ag]SCS-SF: Self-Compassion Scale-Short-Form.
[ah]BESAA: Body Esteem Scale for Adolescents and Adults.
[ai]PACS: Physical Appearance Comparison Scale.
[aj]VAS: visual analog scale.
[ak]SAAS: Social Appearance Anxiety Scale.
[al]UPACS: Upward Physical Appearance Comparison Scale.
[am]BAS-2: Body Appreciation Scale-2.
[an]BIAAQ: Body Image-Acceptance and Action Questionnaire.
[ao]SATAS: Sociocultural Attitudes Toward Appearance Scale.
[ap]WSC: Weight and Shape Concern.
[aq]QDC: Questionario sul Dismorfismo Corporeo.
[ar]DASS-21: Depression Anxiety Stress Scale-21.
[as]EDI-3: Eating Disorder Inventory-3.
[at]SATAQ: Sociocultural AttitudeTowardds Appearance Questionnaire.
[au]BIS-ES: Body Image Self-Efficacy Scale.

XSL•FO
RenderX

Risk of Bias in Studies

Of the 20 included studies, only 4 (20%) were assessed as low risk of bias, while 13 (65%) demonstrated high risk of bias, and 3 (15%) were categorized as moderate risk of bias (see Multimedia Appendix 2).

Random Sequence Generation

This domain assesses whether the method used to generate the random sequence was adequate to prevent selection bias. Although all studies were reported to be RCTs, only 11 studies were assessed as having a low risk of bias in random sequence generation. Computer-generated random numbers were reported in 1 study [46], and a random number table was reported in 3 studies [25,39,42]. Cluster or block randomization was reported in 4 studies [38,41,53,55]. Minimization functions in Qualtrics were reported in 2 studies [40,56]. It was reported in 1 study [54] that whole timeslots were randomly allocated to a condition to avoid cross-contamination in an open computer laboratory, which was assessed as low risk of bias. However, 1 study [44] was assessed as having a high risk of bias in random sequence generation because students were assigned to a group based on class schedule rather than randomization. The remaining 8 studies [43,45,47-52], which did not mention a specific method used to generate the random sequence, were assessed as unclear bias.

Allocation Concealment

This domain evaluates whether the process of assigning participants to groups was concealed. Participants were reported to be randomly allocated to the intervention or control groups in all studies. Only 4 studies [39,41,49,53] reported the use of email to complete allocation concealment, which was assessed as a low risk of bias. In one study [55] reported participants and researchers were not concealed from the randomized arm, which was assessed as a high risk of bias. Due to a lack of further details about allocation concealment, 15 studies [25,38,40,42-48,50-52,54,56] were considered as an unclear risk of bias.

Blinding of Participants and Personnel

This domain assesses whether participants and researchers were blinded to group allocation. It was hard to achieve the blinding of participants or personnel because digital interventions were conducted in intervention groups and nondigital interventions were conducted in control groups in most studies involved. Internet-based or social media assessments and interventions were used in 6 studies [45,46,51,53-55], which may be interpreted as blinding of participants and personnel. Among the 6 studies, 1 study [45] reported that participants used a pseudonym. One study [54] reported that all assessments were self-reported anonymously via computer. One study [39] reported that assessments were housed on a survey software, and the procedure was identical for participants in both the intervention and control groups. Two studies [25,44] mentioned that participants were instructed not to discuss the study with other students to reduce potential cross-contamination, which might be seen as blinding participants and personnel. Four studies [40,42,47,49] were assessed as high risk of bias for their failure to blind participants. Seven studies [38,42,44,46,48,53,56] were assessed as unclear risk of bias due to a lack of further details being reported.

Blinding of Outcome Assessment

This domain evaluates whether the outcome assessors were blinded to the intervention group assignments. Twelve studies were assessed as having a low risk of bias in the blinding of outcome assessment. One [42] of these studies used dummy codes instead of participants' names for outcome assessment. The outcome assessors in another 2 studies [48,55] were unaware of group allocation. Although assessors were not blinded, questionnaires were completed via website, social media, or application by participants in 6 studies [38,39,41,46,53,54]. Another 3 studies [47,49,56] were assessed as having a low risk of bias because questionnaires were sent by e-mail. Eight studies [25,38-40,46-48,50] were assessed as unclear due to the lack of clear information on whether the outcome assessors were blinded to the intervention group assignments. No study was evaluated as having a high risk of bias.

Incomplete Outcome Data

This domain assesses whether there was any missing data and how it was handled. Fourteen studies were assessed to have a low risk of bias. Three [25,41,53] of these studies reported that no missing data emerged. One study [45] reported the usage of baseline measures in place of missing posttreatment or follow-up data. One study [46] claimed that its missing item values were replaced with the mean value of that participant's scale scores. Another 9 studies [38-40,43,47,48,51,54,55] reported low attrition rates with balanced numbers across groups. Four studies [42,44,49,56] were assessed as high risk of bias because of high attrition rates after being randomized. Two studies [50,52] did not report any information about the attrition rate, which was assessed to be at unclear risk of bias.

Selective Reporting

This domain evaluates whether the study reported all the outcomes that were planned at the outset. One study [49] was assessed to be at high risk of bias, as it mentioned that nonsignificant results were not reported. A total of 19 studies [25,38-48,50-56] were assessed as having a low risk of bias.

Other Potential Sources of Bias

This domain assesses other potential biases not covered by the previous domains. Seven studies [25,39,40,42,52,54,55] with microinterventions lasting less than a week were assessed as unclear risk of bias. The rest appeared to be free of other biases (Figure 2) [25,38-45,47-56]. A funnel plot of SE against SMD was generated for 2 outcomes with at least 10 studies: thin-ideal internalization and body dissatisfaction. A visual inspection of the plot for the 2 funnel plots revealed the presence of publication bias (Multimedia Appendix 3).

XSL•FO
RenderX

Figure 2. Risk of bias graph [25,38-45,47-56].

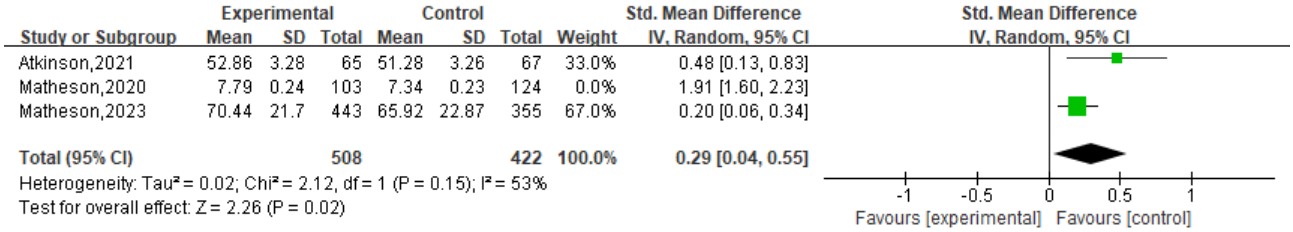
XSL•FO
RenderX

Results of Syntheses

The Effect of Digital Interventions on Body Image Satisfaction or Dissatisfaction

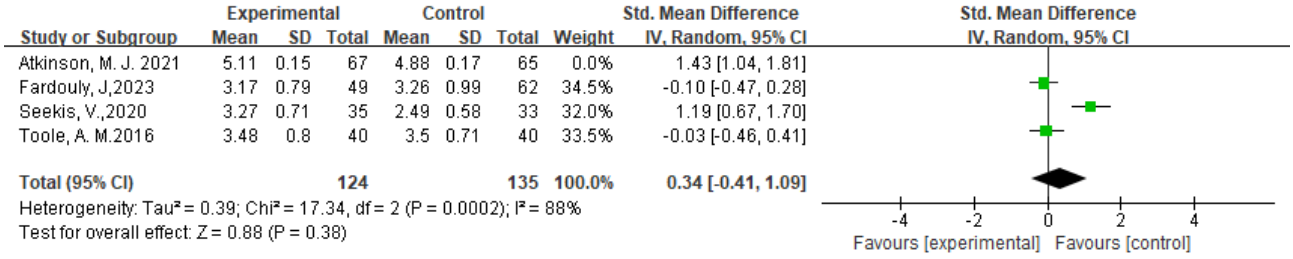
Three studies [40,42,54] reported the original data of body satisfaction among 1157 children and emerging adults by using the body satisfaction subscale of Visual Analogue Scale (VAS), and the higher scores indicated higher body satisfaction. The effect of digital interventions on body satisfaction was shown to be nonsignificant in the random-effect model (SMD=0.86, 95% CI −0.17 to 1.89) with a high level of heterogeneity across studies (I^2=98%; P=.10). This may be due to a diverse rating system, that is, one study used the mean score for each item, while another two studies used a summed score. While the heterogeneity reduced (I^2=53%) after one study [40] was excluded, and the result was a statistically significant conclusion with a small effect size (SMD=0.29, 95% CI 0.04 to 0.55; P=.02; Figure 3 [40,42,54]).

Figure 3. The effect of digital interventions on body satisfaction [40,42,54].



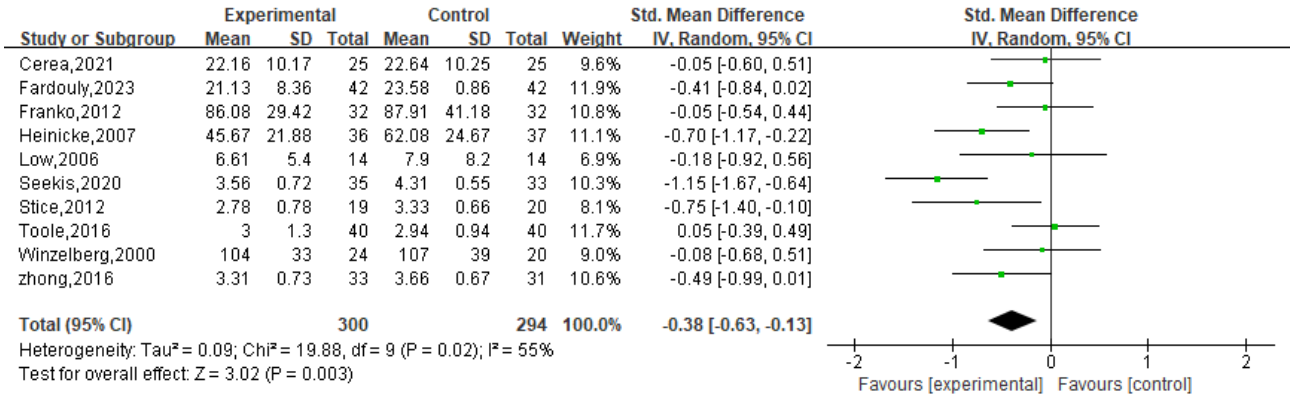
Four studies [51,53,54,56] assessed the effect of digital interventions on body appreciation among 391 children and emerging adults. The pooled analysis showed a nonsignificant improvement in body appreciation between groups in the random-effects model with a high level of heterogeneity (SMD=0.62, 95% CI −0.19 to 1.43; I^2=93%; P=.13). When excluding the study [54] with the outlying effect size, the overall effect did not change (SMD=0.34, 95% CI −0.41 to 1.09; P=.038; Figure 4 [51,53,54,56]), and the heterogeneity remained high (I^2=88%).

Figure 4. The effect of digital interventions on body appreciation [51,53,54,56].



Ten RCTs [41,43,45-48,50,51,53,56] assessed the effect of digital interventions on body dissatisfaction among 594 participants, and the higher scores of the scales used in these studies indicated higher body dissatisfaction. The effect of digital interventions on body dissatisfaction was shown to be statistically significant with a small effect size in the random-effect model (SMD=−0.38, 95% CI: −0.63 to −0.13; I^2=55%; P=.003; Figure 5 [41,43,45-48,50,51,53,56]).

Figure 5. The effect of digital interventions on body dissatisfaction [41,43,45-48,50,51,53,56].

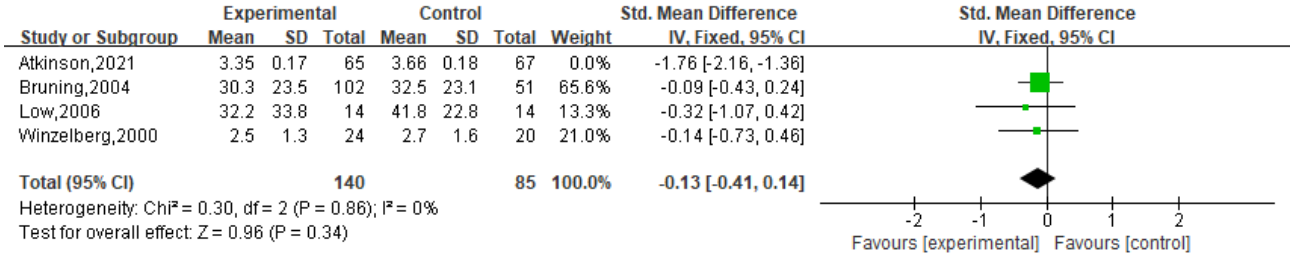


Four RCTs [43-45,54], including 357 participants, reported the effectiveness of digital interventions on the shape and weight concerns. Shape and weight concern is a facet of negative body image, and higher scores on the scales in the included studies reflected greater body dissatisfaction and less body image satisfaction. The pooled analysis showed non-significant improvement (SMD=−0.59, 95% CI −1.5 to 0.32) in shape and weight concern with high heterogeneity (I^2=93%; P=.21). There was no heterogeneity (I^2=0%) after one study [54] was excluded

XSL•FO
RenderX

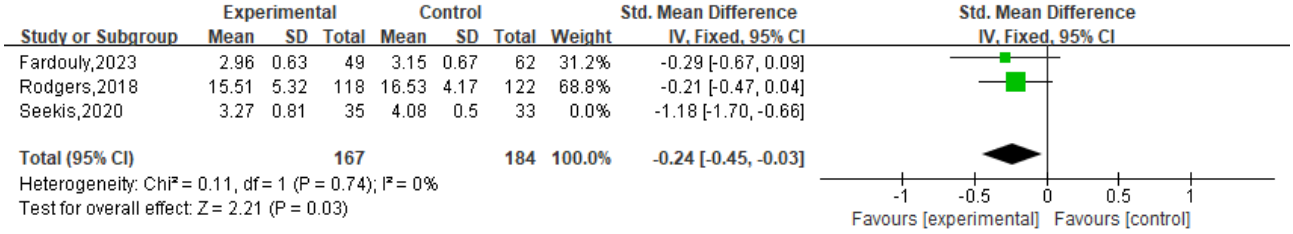
(SMD=−0.13, 95% CI −0.41 to 0.14; *P*=.34; Figure 6 [43-45,54]). However, the result was not significantly different between groups.

Figure 6. The effect of digital interventions on shape and weight concerns [43-45,54].



Three RCTs [39,53,56] measured the effects of digital interventions on physical appearance comparison among 419 participants. The higher the scores of the scales included in the studies, the more severe the physical appearance comparison among the participants. The result showed a statistically significant conclusion with high heterogeneity (SMD=−0.52, 95% CI −1.02 to −0.01; I^2=82%; *P*=.004). The heterogeneity decreased to low with a small effect size (SMD=−0.24, 95% CI −0.45 to −0.03; I^2=0%; *P*=.03; Figure 7 [39,53,56]) when one study [53] was excluded.

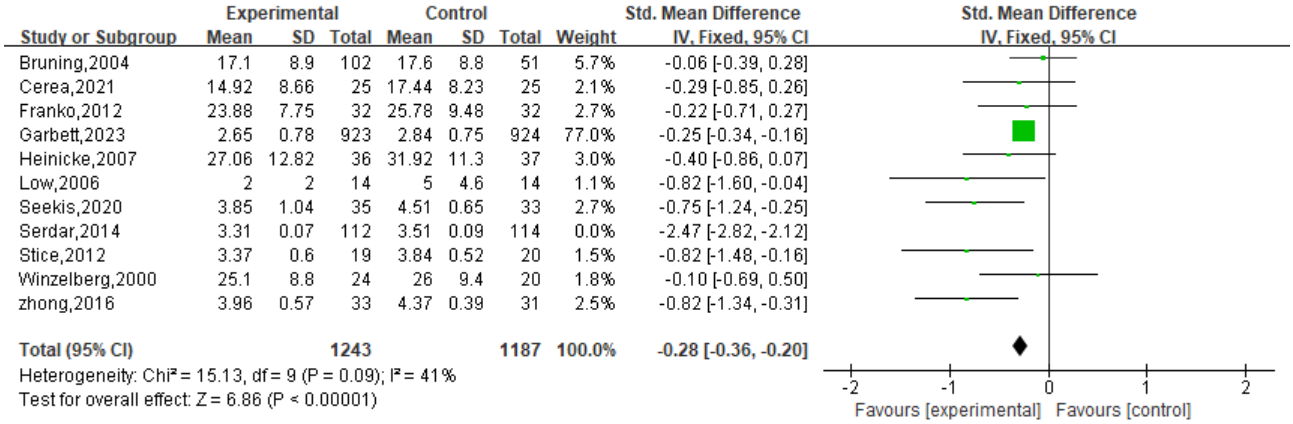
Figure 7. The effect of digital interventions on physical appearance comparison [39,53,56].



Effect of Digital Interventions on Internalized Outcomes

Eleven RCTs [41,43-50,53,55] measured the effect of digital interventions on thin-ideal internalization among 3362 participants. The higher the scores of thin-ideal internalization, the stronger the desire to be thinner, not objectively thin. The pooled analysis showed a significant improvement in thin-ideal internalization in the random-effect model (SMD=−0.64, 95% CI −1.09 to −0.18) with high heterogeneity (I^2=94; *P*<.001). The heterogeneity reduced to moderate (I^2=41%) when one study [49] was excluded, while the result remained a small and significant effect size (SMD=−0.28, 95% CI −0.36 to −0.20; *P*=.001; Figure 8 [41,43-50,53,55]).

Figure 8. The effect of digital interventions on thin-ideal internalization [41,43-50,53,55].



Six RCTs [25,38,39,42,49,55] measured the effectiveness of digital interventions on self-esteem among 3632 participants. Higher scores for self-esteem indicate higher body image satisfaction. The result depicted self-esteem in the experimental group was significantly better than the control group with high heterogeneity (SMD=0.28, 95% CI 0.08-0.47; I^2=81%; *P*=.005). When excluding the study [49], the intervention effect on self-esteem remained significant with a small effect size (SMD=0.14, 95% CI 0.07-0.21) with low heterogeneity (I^2=21%, *P*<.001; Multimedia Appendix 4).

Meta-analysis results on 3 studies [39,51,53] with 385 participants showed that self-compassion in the experimental group was significantly better than the control group with a high level of heterogeneity (SMD=0.73, 95% CI 0.25-1.2; I^2=76%; *P*=.003). The heterogeneity reduced to moderate (I^2=35%) after excluding the study [53], and the intervention effect on

XSL•FO
RenderX

self-compassion remained significant with a medium effect size (SMD=0.55, 95% CI 0.33-0.78; *P*=.001; Multimedia Appendix 5).

Two RCTs [52,56] assessed the effect of digital interventions on self-objectification among 191 participants. Self-objectification refers to conceptualizing one's own body as objects to be scrutinized by others, and higher levels of self-objectification indicate greater body dissatisfaction. The pooled analysis showed that self-objectification in the experimental groups was not significantly different from the control groups (SMD=−0.05, 95% CI −0.33 to 0.24; I^2=0%; *P*=.75; Multimedia Appendix 6).

Effect of Digital Interventions on Negative Affect

Four RCTs [39,42,55,56] assessed the effect of digital interventions on negative affect among 2981 participants. The negative effects include afraid, upset, shame, hostility, and misery, and higher scores of negative affect indicate higher body dissatisfaction. The pooled analysis showed that the negative affect in the experimental groups was not superior to the control groups (SMD=−0.07, 95% CI −0.14 to 0.00; I^2=0%; *P*=.05; Multimedia Appendix 7).

Four RCTs reported the effect of digital interventions on depression, anxiety, and stress among 342 participants by using the Beck Depression Inventory (BDI) [46,48], the Depression Anxiety Stress Scale-21 (DASS-21) [41], and the Social Appearance Anxiety Scale (SAAS) [53]. The higher scores in the 3 scales indicate higher depression, anxiety, and stress. Owing to the assessment of different psychological outcomes, the data could not be synthesized in the 2 studies [41,53], and there was no statistically significant improvement found in the 2 studies, separately. However, the pooled analysis of the BDI group showed a significant improvement in depression in the fixed-effects model with a medium effect size (SMD=−0.59, 95% CI −0.97 to −0.21; I^2=0%; *P*=.002; Multimedia Appendix 8).

Effect of Digital Interventions on Eating Behaviors

Four RCTs [41,44,45,49] assessed the effect of digital interventions on eating disorders among 457 participants. The pooled analysis of eating disorders showed there was no statistically significant difference between groups (SMD=−0.37, 95% CI −1.36 to 0.6) with high heterogeneity (I^2=95%; *P*=.45). The heterogeneity decreased to low (I^2=0%) after one study [49] was excluded, and the result remained no significant difference between groups (SMD=−0.04, 95% CI −0.31 to 0.22; *P*=.75; Multimedia Appendix 9).

Two RCTs [46,48] assessed the effect of digital interventions on eating restraint among 203 participants. Higher scores of eating restraint are indicative of greater use of extreme weight loss behaviors. The pooled analysis of eating restraint showed that there was no statistically significant difference between groups (SMD=−0.36, 95% CI −0.71 to 0.00; I^2=31%; *P*=.05; Multimedia Appendix 10).

Discussion

Principal Findings

This systematic review and meta-analysis synthesized evidence from 20 RCTs involving 5251 participants to evaluate the effectiveness of digital interventions (eg, web pages, mobile apps, and social media) on body dissatisfaction and related outcomes. Key findings indicated that digital interventions significantly improved body dissatisfaction, physical appearance comparison, thin-ideal internalization, self-esteem, self-compassion, and depression, with small to medium effect sizes. However, there was substantial heterogeneity across studies, and potential publication bias was detected. Digital interventions did not significantly improve negative affect (including depression and anxiety) and eating behaviors.

Study Characteristics

All the included studies have been conducted in the last 2 decades, but only 2 studies have been done in developing countries. This indicates that high-quality RCTs need to be conducted further in this area, especially to explore their effectiveness on body dissatisfaction in different ethnic and cultural backgrounds in developing countries [57]. Ten out of twenty included studies selected different theories or models to motivate participants to reduce the pursuit of thin-ideal internalization, to improve emotional regulation, such as self-compassion, and to develop coping strategies with the involvement of parents, peers, and the media. The cognitive dissonance theory, cognitive behavioral theory, and sociocultural theory were recognized as the most widely mentioned theories to reduce body dissatisfaction in the studies. The average sample size was high 200s, but most studies included participants of less than 100. This would result from the narrow age range and the diagnosis of anorexia nervosa and other eating or psychiatric disorders among children and emerging adults in the inclusion criteria. The participants' ages ranged from 7 to 25 years old, and 6 studies included those younger than 14 years old in private or public schools as participants. Nearly 80% of participants were females in this review, and 9 studies only included female college students or undergraduates as participants, which accounted for 16.7% of the total participants. This can indicate that females would be more likely to experience body dissatisfaction, like other studies [58].

Comparison to Previous Work

The meta-analysis revealed that digital interventions showed a small to medium effect across most outcomes in body satisfaction and internalized outcomes, and the result was similar to another systematic review [24]. In the systematic review, although there was no quantitative synthesis, body dissatisfaction, body esteem, body appreciation, and other body image outcomes were also included in the review to explore the effectiveness of digital interventions in emerging adults and emerging adults. The results showed 8 out of 15 studies reported digital interventions were effective in improving at least one body image outcome from pre-post interventions with mostly small to medium effect sizes. This means digital interventions are superior to some traditional interventions, such as brochures, documentaries, newspapers, or magazines. It may result from

XSL•FO
RenderX

the advantages of digital interventions, such as real-time feedback and motivation enhancement [59], and anonymity of platforms [60], which could reduce the shame and social anxiety emerging adults face when dealing with body image-related issues. However, given that effect sizes were small to medium, it would be caused by the following reasons. First, the intervention duration of this meta-analysis was 6 weeks on average. Second, digital intervention had its limitations, such as being incapable of genuinely caring about one's feelings during the intervention process. This indicated that digital interventions with an in-person element would be associated with greater effectiveness in body image [61].

Digital interventions did not significantly reduce the negative affect (including depression and anxiety) and eating behaviors. However, another 2 systematic reviews [61,62] found a significant effect of digital interventions on depression and anxiety among emerging adults and emerging adults. This would result from the contents of the digital interventions. The digital interventions of the 2 systematic reviews specifically targeted mental health disorders, including depression and anxiety. However, this meta-analysis mainly targeted improving body dissatisfaction with depression and anxiety as secondary outcomes in some studies included. Furthermore, it indicated that digital psychotherapy did not significantly reduce both eating disorders and restrained eating in this meta-analysis. This would be caused by the following reasons. First, no more than 4 RCTs were included in this meta-analysis to explore the effects of digital interventions on eating disorders or restraint eating. Second, the age of participants in this meta-analysis ranged from 7 to 25 years. Especially for the emerging adults who were predominant participants, who have already established eating behavior habits, which would be so ingrained and resistant to change through short-term digital psychological therapy. According to the Transtheoretical Model [63], the process of behavior change is complicated with 6 different stages, and it would not be easy to change eating behaviors. In addition, the digital interventions in the 20 included RCTs were aimed at addressing body dissatisfaction rather than specific eating disorder symptoms, which might account for the insignificant outcome in changing eating behaviors.

Implications for Clinical Practice

The findings suggest that digital interventions can serve as scalable and cost-effective supplementary tools to improve body dissatisfaction in children and emerging adults. Clinicians may recommend digital interventions as adjuncts to traditional face-to-face therapies, particularly for individuals with limited access to in-person mental health services due to cost, stigma, or geographic barriers. Body dissatisfaction, thin-ideal internalization, eating disorders, and negative affect should be addressed simultaneously in a single intervention to verify the effectiveness of digital interventions in the future. Future studies should be conducted in Asia and some other low-income countries, as this review found that studies in these regions are lacking. Furthermore, the quality of included studies held room for improvement.

Limitations

This systematic review has several limitations. First, the high heterogeneity in meta-analysis may result from the use of different scales to assess the same outcomes, even using different versions or varying score systems (eg, summed versus mean scores) of the same scales. However, the SMD was selected to express the size of the intervention effect, instead of the mean difference. Second, confidence in the meta-analysis was limited because waitlist groups were set in 7 studies, and assessment-only control (receive no intervention) groups were set in another 4 studies, rather than active control groups. Third, there was insufficient data on long‐term follow‐up and a potential gap that future RCTs can look to fill. Specific tactics could be analyzed to maintain engagement in body image interventions for children and emerging adults, such as integrating digital interventions into existing curricula. Fourth, the quality of included studies held room for improvement, with most of the RCTs included being rated as "unclear risk" for bias. Fifth, publication bias may have influenced the results, particularly for outcomes with small sample sizes, highlighting the need for larger, registered RCTs.

Conclusion

Digital interventions could help children and emerging adults improve body satisfaction or dissatisfaction, physical appearance comparison, thin-ideal internalization, self-esteem, self-compassion, and depression. Because of the limitations, the results should be generalized with caution. In the future, high-quality RCTs with longer intervention duration and long-term follow-up should be conducted, especially in different races and cultures in transitional countries.

Acknowledgments

The authors would like to thank Professor Roger Watson for his help with the review statement and English language editing.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

LL: Writing – original draft, Conceptualization, Data curation, Methodology, Software. JY: Writing – original draft, Conceptualization, Data curation. FT: Conceptualization, Data curation, Methodology. XY: Conceptualization, Data curation,

XSL•FO
RenderX

Methodology. HL: Conceptualization, Methodology. YC: Supervision, Writing – review & editing. XLZ: Supervision, Writing – review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOC File, 31 KB - i-jmr_v14i1e72231_app1.doc]

Multimedia Appendix 2
Study quality assessment.
[DOC File, 73 KB - i-jmr_v14i1e72231_app2.doc]

Multimedia Appendix 3
Funnel plot.
[DOC File, 35 KB - i-jmr_v14i1e72231_app3.doc]

Multimedia Appendix 4
The effect of digital interventions on self-esteem [25,38,39,42,49,55].
[DOC File, 33 KB - i-jmr_v14i1e72231_app4.doc]

Multimedia Appendix 5
The effect of digital interventions on self-compassion [39,51,53].
[DOC File, 31 KB - i-jmr_v14i1e72231_app5.doc]

Multimedia Appendix 6
The effect of digital interventions on self-objectification [52,56].
[DOC File, 30 KB - i-jmr_v14i1e72231_app6.doc]

Multimedia Appendix 7
The effect of digital interventions on negative affect [39,42,55,56].
[DOC File, 32 KB - i-jmr_v14i1e72231_app7.doc]

Multimedia Appendix 8
The effect of digital interventions on depression [41,53].
[DOC File, 30 KB - i-jmr_v14i1e72231_app8.doc]

Multimedia Appendix 9
The effect of digital interventions on eating disorders [41,44,45,49].
[DOC File, 32 KB - i-jmr_v14i1e72231_app9.doc]

Multimedia Appendix 10
The effect of digital interventions on eating restraint [46,48].
[DOC File, 30 KB - i-jmr_v14i1e72231_app10.doc]

Checklist 1
PRISMA 2020 checklist.
[DOC File, 335 KB - i-jmr_v14i1e72231_app11.doc]

References

1. Himmerich H, Mirzaei K. Body image, nutrition, and mental health. Nutrients 2024 Apr 10;16(8):1106. [doi: 10.3390/nu16081106] [Medline: 38674797]
2. Yuan Y, Hu J, Sun L, et al. An association between body image dissatisfaction and digit ratio among Chinese children and adolescents. Sci Rep 2021 Mar 4;11(1):5217. [doi: 10.1038/s41598-021-84711-x] [Medline: 33664410]

XSL•FO
RenderX

3. Solano-Pinto N, Sevilla-Vera Y, Fernández-Cézar R, Garrido D. Can parental body dissatisfaction predict that of children? A study on body dissatisfaction, body mass index, and desire to diet in children aged 9-11 and their families. Front Psychol 2021;12:650744. [doi: 10.3389/fpsyg.2021.650744] [Medline: 33868129]
4. Divecha CA, Simon MA, Asaad AA, Tayyab H. Body image perceptions and body image dissatisfaction among medical students in Oman. Sultan Qaboos Univ Med J 2022 May;22(2):218-224. [doi: 10.18295/squmj.8.2021.121] [Medline: 35673292]
5. Gonzaga I, Ribovski M, Claumann GS, et al. Secular trends in body image dissatisfaction and associated factors among adolescents (2007-2017/2018). PLoS One 2023;18(1):e0280520. [doi: 10.1371/journal.pone.0280520] [Medline: 36656894]
6. Liyanage G, Karunainathan T, Jeyarajah L, Thevatheepan P, Thavendra M, Seneviwickrama M. Body image dissatisfaction and its determinants in urban Sri Lankan adolescents. Ceylon Med J 2021 Dec 31;66(4):185-190. [doi: 10.4038/cmj.v66i4.9509] [Medline: 35570350]
7. Alharballeh S, Dodeen H. Prevalence of body image dissatisfaction among youth in the United Arab Emirates: gender, age, and body mass index differences. Curr Psychol 2023;42(2):1317-1326. [doi: 10.1007/s12144-021-01551-8] [Medline: 33679115]
8. Hicks RE, Kenny B, Stevenson S, Vanstone DM. Risk factors in body image dissatisfaction: gender, maladaptive perfectionism, and psychological wellbeing. Heliyon 2022 Jun;8(6):e09745. [doi: 10.1016/j.heliyon.2022.e09745] [Medline: 35770148]
9. Kuck N, Cafitz L, Bürkner PC, Hoppen L, Wilhelm S, Buhlmann U. Body dysmorphic disorder and self-esteem: a meta-analysis. BMC Psychiatry 2021 Jun 15;21(1):310. [doi: 10.1186/s12888-021-03185-3] [Medline: 34130638]
10. Vannucci A, Ohannessian CM. Body image dissatisfaction and anxiety trajectories during adolescence. J Clin Child Adolesc Psychol 2018;47(5):785-795. [doi: 10.1080/15374416.2017.1390755] [Medline: 29087230]
11. Blundell E, De Stavola BL, Kellock MD, et al. Longitudinal pathways between childhood BMI, body dissatisfaction, and adolescent depression: an observational study using the UK Millenium Cohort Study. Lancet Psychiatry 2024 Jan;11(1):47-55. [doi: 10.1016/S2215-0366(23)00365-6] [Medline: 38101872]
12. Foster L, Lundh LG, Daukantaité D. Disordered eating in a 10-year perspective from adolescence to young adulthood: Stability, change, and body dissatisfaction as a predictor. Scand J Psychol 2024 Feb;65(1):32-41. [doi: 10.1111/sjop.12950] [Medline: 37491950]
13. Neumark-Sztainer D, Paxton SJ, Hannan PJ, Haines J, Story M. Does body satisfaction matter? Five-year longitudinal associations between body satisfaction and health behaviors in adolescent females and males. J Adolesc Health 2006 Aug;39(2):244-251. [doi: 10.1016/j.jadohealth.2005.12.001] [Medline: 16857537]
14. Milton A, Hambleton A, Roberts A, et al. Body image distress and Its associations from an international sample of men and women across the adult life span: web-based survey study. JMIR Form Res 2021 Nov 4;5(11):e25329. [doi: 10.2196/25329] [Medline: 34734831]
15. Sharp G, Tiggemann M, Mattiske J. The role of media and peer influences in Australian women's attitudes towards cosmetic surgery. Body Image 2014 Sep;11(4):482-487. [doi: 10.1016/j.bodyim.2014.07.009] [Medline: 25129686]
16. Yetsenga R, Banerjee R, Streatfeild J, et al. The economic and social costs of body dissatisfaction and appearance-based discrimination in the United States. Eat Disord 2024;32(6):572-602. [doi: 10.1080/10640266.2024.2328461] [Medline: 38520696]
17. Zamiri-Miandoab N, Hassanzadeh R, Kamalifard M, Mirghafourvand M. The effect of cognitive behavior therapy on body image and self-esteem in female adolescents: a systematic review and meta-analysis. Int J Adolesc Med Health 2021 Apr 30;33(6):323-332. [doi: 10.1515/ijamh-2021-0029] [Medline: 33915608]
18. Ciao AC, Munson BR, Pringle KD, et al. Inclusive dissonance-based body image interventions for college students: Two randomized-controlled trials of the EVERYbody Project. J Consult Clin Psychol 2021 Apr;89(4):301-315. [doi: 10.1037/ccp0000636] [Medline: 34014692]
19. Pursey KM, Burrows TL, Barker D, Hart M, Paxton SJ. Disordered eating, body image concerns, and weight control behaviors in primary school aged children: A systematic review and meta-analysis of universal-selective prevention interventions. Int J Eat Disord 2021 Oct;54(10):1730-1765. [doi: 10.1002/eat.23571] [Medline: 34245459]
20. Doley JR, Rodgers RF, Paxton SJ, McLean SA. Effectiveness of recruitment strategies for a social media literacy E-intervention for young adults with body dissatisfaction: cost, time, diversity, and completion. Eat Disord 2022;30(5):515-539. [doi: 10.1080/10640266.2021.1938855] [Medline: 34402410]
21. Goldstein KM, Rushton S, Lewinski AA, et al. Experiences receiving and delivering virtual health care for women: qualitative evidence synthesis. J Med Internet Res 2025 May 15;27:e68314. [doi: 10.2196/68314] [Medline: 40373303]
22. Rounsefell K, Gibson S, McLean S, et al. Social media, body image and food choices in healthy young adults: A mixed methods systematic review. Nutr Diet 2020 Feb;77(1):19-40. [doi: 10.1111/1747-0080.12581] [Medline: 31583837]
23. Plackett R, Steward JM, Kassianos AP, et al. The effectiveness of social media campaigns in improving knowledge and attitudes toward mental health and help-seeking in high-income countries: scoping review. J Med Internet Res 2025 May 23;27:e68124. [doi: 10.2196/68124] [Medline: 40408767]
24. Mahon C, Seekis V. Systematic review of digital interventions for adolescent and young adult women's body image. Front Glob Womens Health 2022;3:832805. [doi: 10.3389/fgwh.2022.832805] [Medline: 35392118]

XSL•FO
RenderX

25. Halliwell E, Easun A, Harcourt D. Body dissatisfaction: can a short media literacy message reduce negative media exposure effects amongst adolescent girls? Br J Health Psychol 2011 May;16(Pt 2):396-403. [doi: 10.1348/135910710X515714] [Medline: 21489065]
26. 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]
27. Convention on the rights of the child: the children's version. Unicef. 2025. URL: https://www.unicef.org/child-rights-convention/convention-text-childrens-version [accessed 2025-07-15]
28. Leung TS, Zeng G, Maylott SE, et al. Emerging adults mirror infants' emotions and yawns. Dev Psychobiol 2024 Sep;66(6):e22539. [doi: 10.1002/dev.22539] [Medline: 39164829]
29. Thompson JK, Coovert MD, Stormer SM. Body image, social comparison, and eating disturbance: a covariance structure modeling investigation. Int J Eat Disord 1999 Jul;26(1):43-51. [doi: 10.1002/(sici)1098-108x(199907)26:1<43::aid-eat6>3.0.co;2-r] [Medline: 10349583]
30. Paterna A, Alcaraz-Ibáñez M, Fuller-Tyszkiewicz M, Sicilia Á. Internalization of body shape ideals and body dissatisfaction: a systematic review and meta-analysis. Int J Eat Disord 2021 Sep;54(9):1575-1600. [doi: 10.1002/eat.23568] [Medline: 34165215]
31. Sharpe H, Griffiths S, Choo TH, et al. The relative importance of dissatisfaction, overvaluation and preoccupation with weight and shape for predicting onset of disordered eating behaviors and depressive symptoms over 15 years. Int J Eat Disord 2018 Oct;51(10):1168-1175. [doi: 10.1002/eat.22936] [Medline: 30194690]
32. Stice E, Shaw HE. Role of body dissatisfaction in the onset and maintenance of eating pathology: a synthesis of research findings. J Psychosom Res 2002 Nov;53(5):985-993. [doi: 10.1016/s0022-3999(02)00488-9] [Medline: 12445588]
33. Higgins JPT, Altman DG, Gøtzsche PC, et al. The Cochrane collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343(oct18 2):d5928. [doi: 10.1136/bmj.d5928] [Medline: 22008217]
34. Claudino JG, Gabbett TJ, Bourgeois F, et al. CrossFit overview: systematic review and meta-analysis. Sports Med Open 2018 Feb 26;4(1):11. [doi: 10.1186/s40798-018-0124-5] [Medline: 29484512]
35. Muller K. Statistical power analysis for the behavioral sciences. Technometrics 1989 Nov;31(4):499-500. [doi: 10.1080/00401706.1989.10488618]
36. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 6;327(7414):557-560. [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
37. Tchero H, Tabue Teguo M, Lannuzel A, Rusch E. Telerehabilitation for stroke survivors: systematic review and meta-analysis. J Med Internet Res 2018 Oct 26;20(10):e10867. [doi: 10.2196/10867] [Medline: 30368437]
38. Cousineau TM, Franko DL, Trant M, et al. Teaching adolescents about changing bodies: randomized controlled trial of an Internet puberty education and body dissatisfaction prevention program. Body Image 2010 Sep;7(4):296-300. [doi: 10.1016/j.bodyim.2010.06.003] [Medline: 20638919]
39. Rodgers RF, Donovan E, Cousineau T, et al. BodiMojo: efficacy of a mobile-based intervention in improving body image and self-compassion among adolescents. J Youth Adolesc 2018 Jul;47(7):1363-1372. [doi: 10.1007/s10964-017-0804-3] [Medline: 29349593]
40. Matheson EL, Lewis-Smith H, Diedrichs PC. The effectiveness of brief animated films as a scalable micro-intervention to improve children's body image: A randomised controlled trial. Body Image 2020 Dec;35:142-153. [doi: 10.1016/j.bodyim.2020.08.015] [Medline: 33049455]
41. Cerea S, Ghisi M, Bottesi G, Manoli T, Carraro E, Doron G. Cognitive behavioral training using a mobile application reduces body image-related symptoms in high-risk female University students: a randomized controlled study. Behav Ther 2021 Jan;52(1):170-182. [doi: 10.1016/j.beth.2020.04.002] [Medline: 33483115]
42. Matheson EL, Smith HG, Amaral ACS, et al. Using chatbot technology to improve Brazilian adolescents' body image and mental health at scale: randomized controlled trial. JMIR Mhealth Uhealth 2023 Jun 19;11:e39934. [doi: 10.2196/39934] [Medline: 37335604]
43. Winzelberg AJ, Eppstein D, Eldredge KL, et al. Effectiveness of an internet-based program for reducing risk factors for eating disorders. J Consult Clin Psychol 2000 Apr;68(2):346-350. [doi: 10.1037//0022-006x.68.2.346] [Medline: 10780136]
44. Bruning Brown J, Winzelberg AJ, Abascal LB, Taylor CB. An evaluation of an Internet-delivered eating disorder prevention program for adolescents and their parents. J Adolesc Health 2004 Oct;35(4):290-296. [doi: 10.1016/j.jadohealth.2003.10.010] [Medline: 15450542]
45. Low KG, Charanasomboon S, Lesser J, et al. Effectiveness of a computer-based interactive eating disorders prevention program at long-term follow-up. Eat Disord 2006;14(1):17-30. [doi: 10.1080/10640260500403816] [Medline: 16757446]
46. Heinicke BE, Paxton SJ, McLean SA, Wertheim EH. Internet-delivered targeted group intervention for body dissatisfaction and disordered eating in adolescent girls: a randomized controlled trial. J Abnorm Child Psychol 2007 Jun;35(3):379-391. [doi: 10.1007/s10802-006-9097-9] [Medline: 17243014]
47. Franko DL, Jenkins A, Rodgers RF. Toward reducing risk for eating disorders and obesity in Latina College Women. J Couns Dev 2012 Jul;90(3):298-307 [FREE Full text] [doi: 10.1002/j.1556-6676.2012.00038.x]

XSL•FO
RenderX

48. Stice E, Rohde P, Durant S, Shaw H. A preliminary trial of a prototype Internet dissonance-based eating disorder prevention program for young women with body image concerns. J Consult Clin Psychol 2012 Oct;80(5):907-916. [doi: 10.1037/a0028016] [Medline: 22506791]
49. Serdar K, Kelly NR, Palmberg AA, et al. Comparing online and face-to-face dissonance-based eating disorder prevention. Eat Disord 2014;22(3):244-260. [doi: 10.1080/10640266.2013.874824] [Medline: 24456277]
50. Yunyu Z. Towards evaluating an internet-based dissonance intervention to reduce body dissatisfaction among young Chinese women: some preliminary findings. Community Psychol Res 2016;2:150-167 [FREE Full text]
51. Toole AM, Craighead LW. Brief self-compassion meditation training for body image distress in young adult women. Body Image 2016 Dec;19:104-112. [doi: 10.1016/j.bodyim.2016.09.001] [Medline: 27664531]
52. Slater A, Halliwell E, Jarman H, Gaskin E. More than Just Child's Play?: An experimental investigation of the impact of an appearance-focused internet game on body image and career aspirations of young girls. J Youth Adolesc 2017 Sep;46(9):2047-2059. [doi: 10.1007/s10964-017-0659-7] [Medline: 28316057]
53. Seekis V, Bradley GL, Duffy AL. Does a Facebook-enhanced mindful self-compassion intervention improve body image? An evaluation study. Body Image 2020 Sep;34:259-269. [doi: 10.1016/j.bodyim.2020.07.006] [Medline: 32717627]
54. Atkinson MJ, Diedrichs PC. Examining the efficacy of video-based microinterventions for improving risk and protective factors for disordered eating among young adult women. Int J Eat Disord 2021 May;54(5):708-720. [doi: 10.1002/eat.23460] [Medline: 33415790]
55. Garbett KM, Haywood S, Craddock N, et al. Evaluating the efficacy of a social media-based intervention (Warna-Warni Waktu) to improve body image among young Indonesian women: parallel randomized controlled trial. J Med Internet Res 2023 Apr 3;25:e42499. [doi: 10.2196/42499] [Medline: 37010911]
56. Fardouly J, Slater A, Parnell J, Diedrichs PC. Can following body positive or appearance neutral Facebook pages improve young women's body image and mood? Testing novel social media micro-interventions. Body Image 2023 Mar;44:136-147. [doi: 10.1016/j.bodyim.2022.12.008] [Medline: 36608433]
57. Frederick DA, Murray SB. The U.S. body project I: inspirations and lessons learned. Body Image 2023 Sep;46:223-229. [doi: 10.1016/j.bodyim.2023.05.010] [Medline: 37354878]
58. Huguenin FM, de Almeida VAR, Rodrigues MVF, Ferreira MEC, Morgado FFDR. Body image of university students: a systematic review of the characteristics of interventions. Psicol Reflex Crit 2024 Jun 24;37(1):22. [doi: 10.1186/s41155-024-00307-0] [Medline: 38913140]
59. Frank E, Wallace ML, Matthews MJ, et al. Personalized digital intervention for depression based on social rhythm principles adds significantly to outpatient treatment. Front Digit Health 2022;4:870522. [doi: 10.3389/fdgth.2022.870522] [Medline: 36120713]
60. Rasing SPA, Stikkelbroek YAJ, Bodden DHM. Is digital treatment the holy grail? Literature review on computerized and blended treatment for depressive disorders in youth. Int J Environ Res Public Health 2019 Dec 24;17(1):153. [doi: 10.3390/ijerph17010153] [Medline: 31878249]
61. Lehtimaki S, Martic J, Wahl B, Foster KT, Schwalbe N. Evidence on Digital Mental Health Interventions for Adolescents and Young People: Systematic Overview. JMIR Ment Health 2021 Apr 29;8(4):e25847. [doi: 10.2196/25847] [Medline: 33913817]
62. Christ C, Schouten MJ, Blankers M, et al. Internet and Computer-Based Cognitive Behavioral Therapy for Anxiety and Depression in Adolescents and Young Adults: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Sep 25;22(9):e17831. [doi: 10.2196/17831] [Medline: 32673212]
63. Kleis RR, Hoch MC, Hogg-Graham R, Hoch JM. The Effectiveness of the Transtheoretical Model to Improve Physical Activity in Healthy Adults: A Systematic Review. J Phys Act Health 2021 Jan 1;18(1):94-108. [doi: 10.1123/jpah.2020-0334] [Medline: 33260143]

Abbreviations

BDI: Beck Depression Inventory
BIS: behavior interventions and supports
CBT: cognitive-behavioral therapy
CNKI: China National Knowledge Infrastructure
DASS-21: Depression Anxiety Stress Scale-21
EBSCO: Elton B. Stephens Company
PICOS: population, intervention, comparison, outcome, and study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SAAS: Social Appearance Anxiety Scale
SMD: standardized mean difference
VAS: Visual Analogue Scale

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 06.02.25; peer-reviewed by NJ Goel, P Shi; revised version received 22.06.25; accepted 23.06.25; published 13.08.25.

Please cite as:
Liu L, Yang J, Tan F, Yang X, Luo H, Chen Y, Zhao X
Digital Interventions for Improving Body Dissatisfaction in Children and Emerging Adults: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e72231
URL: https://www.i-jmr.org/2025/1/e72231
doi:10.2196/72231

© Li Liu, Jianning Yang, Fengmei Tan, Xia Yang, Huan Luo, Yanhua Chen, Xiaolei Zhao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 13.8.2025. 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

Horizon Scanning Methods for Health Care Technology Innovation Identification: Rapid Scoping Review of Patent Research Studies

Sonia Garcia Gonzalez-Moral[1*], MSc; Erin Pennock[2*], MSc; Olushola Ewedairo[1*], MSc; Elizabeth Green[1*], MSc; James Elgey[1*], MSc; Andrew Mkwashi[1*], PhD

[1]NIHR Innovation Observatory, Population Health Sciences Institute, Newcastle University, The Catalyst Room 3.12, 3 Science Square Newcastle Helix, Newcastle upon Tyne, United Kingdom

[2]Center for Cancer, Population Health Sciences Insitute, Newcastle University, Newcastle upon Tyne, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Sonia Garcia Gonzalez-Moral, MSc
NIHR Innovation Observatory, Population Health Sciences Institute, Newcastle University, The Catalyst Room 3.12, 3 Science Square Newcastle Helix, Newcastle upon Tyne, United Kingdom

Abstract

Background: Patents are an early sign of innovation, yet their role in horizon scanning for health care remains unclear.

Objective: This study investigates the role of, and methods for, patent analysis in advancing health care technology innovation in a sector that is characterized by diverse health care technologies and significant research investment. Patents are critical early indicators of innovation, supporting horizon scanning and weak signal detection. The study aimed to identify intellectual property sources, evaluate methods for patent retrieval and analysis, and outline objectives for using patent data to anticipate trends and inform health care strategies.

Methods: A rapid scoping review was conducted following Cochrane Rapid Review Methods recommendations and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, with a preregistered protocol on the Open Science Framework. Searches in Embase, IEEE Xplore, and Web of Science targeted records published 2020 onward to capture the most recent sources, methods, and tools. Three independent reviewers screened studies using Rayyan (Qatar Computing Research Institute). We included any study type published since 2020 that provided patent sources data, methods, and tools applied to the study of health care technologies. Our data extraction included bibliographic details, study characteristics, and methodological information. Risk of bias assessments were not undertaken. Narrative and tabular methods, supplemented by visual charts, were used to synthesize findings.

Results: Our searches identified 1741 studies, of which 124 were included after title, abstract, and full-text screening, with 54% being original research, 43.5% reviews, and the remainder being conference abstracts (2.5%). Most studies (68%) relied solely on patent databases, while others searched the gray and published literature. Research objectives of the included studies were grouped into 10 themes, with trend analysis (50%) and the provision of recommendations for future research, policy, and strategy development (20%) being the most common. Our review identified up to 47 patent databases, with 27% of studies using multiple sources. Whenever time limits were reported, the mean time horizon for patent searches was 24.6 years, ranging from 1900 to 2019. Automated approaches, used in 33% (n=43) of studies, frequently used tools such as Gephi (Gephi Consortium) for network visualization. Disease mapping based on National Institute for Health and Care Excellence classification indicated that cancer (19%) and respiratory conditions (16%), particularly COVID-19, were key areas.

Conclusions: Patent data are valuable for identifying technological trends and informing policy and research strategies. While patents provide crucial insights into emerging technologies, inconsistent deduplication practices across studies pose the risk of data inflation, accentuating the need for transparency and rigor. Finally, this review emphasized the importance of data transformation and visualization in detecting emerging trends, with Python and R being the most commonly used programming languages for developing custom tools.

(Interact J Med Res 2025;14:e70323) doi:10.2196/70323

KEYWORDS

horizon scanning; patent analysis; medtech; rapid scoping review; health care innovation

XSL•FO
RenderX

Introduction

According to the World Intellectual Property Office, a patent is an exclusive right granted for an invention [1]. From the moment a patent is filed, a full disclosure of the invention to a patent office would have occurred [1]. This information will be made available to the public in due course and often prior to the patent being granted [2]. In the health care technology context, patented innovations may include new pharmaceuticals, diagnostics, and medical devices. Often, inventors will apply for patents before clinical trials are initiated to protect the intellectual property of the innovative technology in the trial. However, in the case of low-risk medical devices, where clinical trials are not always conducted, patents may be one of the few indications of a new innovative product before it is introduced to the market. Therefore, patents may be considered as one of the first signs of innovation and are often searched in horizon scanning studies alone or in combination with other sources of weak signal detection [3].

However, with the complexity of some types of health care technologies such as medical devices, as well as the advent of artificial intelligence and its integration into health care technologies, it is quite difficult to identify early signs of innovation in a timely manner. Some digital applications, for instance, evolve rather rapidly, while most class 1 (low risk) medical devices do not undergo clinical trials. Yet, these technologies serve as a primary source of weak signal detection for emerging health care technologies used by the Innovation Observatory (IO), the UK national horizon scanning and research intelligence body, funded by the National Institute for Health and Care Research [4].

Furthermore, a thorough understanding of the health care technology development pathway from ideation to market is imperative to support early warning and alert systems which will, in turn, help inform strategic decision-making, the formulation of health policy, the allocation of research and development funds, and play a vital role in anticipating regulatory challenges associated with emerging technologies. Despite previous horizon scanning and forecast studies conducted in the context of the UK health care horizon scanning needs, it remains unclear how patent analysis methods are being used in forecasting technology emergence globally and how and whether current horizon scanning methods can be enhanced by the regular use of patent analysis studies [5].

Anticipating health care technology innovation before market entry is a key challenge for horizon scanning activities. However, the role of patent landscape analysis within existing horizon scanning methods remains underexplored. For instance, could routine patent analysis help identify emerging MedTech opportunities earlier, thereby enabling more proactive innovation support? Might it also inform the development of regulatory frameworks to reduce delays in technology adoption? To assess the potential of integrating patent analysis into existing horizon scanning practices, we reviewed recent literature on patent studies related to health care technologies. The study had 3 main aims: to identify key intellectual property sources, such as patent databases; to examine methods used for patent retrieval and analysis; and to determine the primary purposes for using patent analysis in this context. The findings aim to support the standardization of weak signal detection and offer a methodological foundation for integrating patent landscape analysis [6] into horizon scanning, both within the IO and among wider national and international bodies, particularly for early-stage technologies.

Methods

Overview

A rapid scoping review was undertaken to efficiently examine the published literature and answer the predetermined research questions on methods, tools, and purpose for patent landscape analysis [7]. We used a scoping review approach to explore the breadth of the subject area and to map the recent patent analysis methods research landscape, while also accelerating some of the traditionally more time-consuming aspects of a systematic review by embracing the recommendations for rapid reviews set out by the Cochrane Rapid Review Methods working group [8]. The use of rapid methods applied to scoping reviews has been documented as an increasing trend in big umbrella reviews due to the efficiency gain [9,10]. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews for the reporting of items relevant to this review. This review did not quality appraise the included studies as this was considered out of scope for the objectives of this project. A preagreed protocol outlining the methods and objectives of this rapid scoping review was registered in the Open Science Framework in June 2024 [7].

Inclusion Criteria

To outline the inclusion criteria for this rapid scoping review, we followed the Cochrane protocol guide. The review was structured based on the SDMO framework (study type, data, methods, outcomes) [11]. Studies eligible for inclusion needed to meet the following criteria: (1) any type of study or conference abstract in English language, (2) data related to the health care technology were reported, (3) methods and tools used for the analysis were mentioned, and (4) all reported outcomes were considered for inclusion although they were not deemed mandatory for the inclusion of studies in this review in line with the recommendation from Higgins et al [11].

Search Strategy

Following the recommendation for rapid searches, a targeted search strategy was devised and run in Embase (Ovid), which contains MEDLINE, by an experienced information specialist and checked against the PRESS checklist [12]. Our search strategy consisted of subject heading terms in combination with text words in title and abstract. Given the specificity of the topic and to maximize the sensitivity of our searches, we focused on the Method concept (Patent) from the SDMO framework outlined above. This concept was searched in combination with terms for the different types of studies such as “trend,” “analysis,” “landscape,” “mining” or more broadly “research.” This search was then translated and run in 2 more databases selected for their comprehensive coverage of technological and scientific topics: IEEE Xplore digital library and Web of

XSL•FO
RenderX

Science. We searched in the title, abstract, and keyword fields. Time limits were imposed to retrieve records published since 2020 to capture the most contemporary tools and sources. Due to the lack of current guidelines for the application of time limits in searches for rapid reviews, this was a pragmatic decision informed by the recently updated Cochrane Rapid Review Methods guidance and agreed with the information specialist [8]. This approach ensured that our review remained focused on the latest developments and advancements in the field, providing the most relevant and up-to-date information. No language limit or any other limits were used. Records were downloaded into Endnote 20 (Clarivate Analytics) for deduplication. Full search strategies with results are provided in Multimedia Appendix 1.

Data Collection

Deduplicated results were single screened at the title and abstract stage by 3 independent reviewers (SGGM, EP, and OE) separately using Rayyan (Qatar Computing Research Institute [QCRI]), a screening software for systematic reviews [13]. Consultation between reviewers was practiced at this stage on multiple occasions. In case of doubt due to ambiguity of titles or lack of clear reporting in abstracts, records were included for full-text screening. At full-text screening, the same 3 reviewers proceeded independently to single screen the full texts for inclusion. One-to-one consultation was exercised regularly throughout this process, and disagreements were resolved by checking with a fourth reviewer.

Data Extraction

To ensure data extraction followed the recommendations of rapid review methods, an abbreviated data extraction form was developed to meet the agreed objectives of this review and was reported in the published protocol [7]. The original data extraction form was then piloted by the reviewers testing the relevance of data fields deemed for extraction. This step did not result in further alterations to the original data extraction form. Three independent reviewers extracted data, and one independently assessed its quality. We extracted bibliographic details; study-related characteristics such as the type of technologies, health condition, and objectives of the study; methods-related data including the sources searched, the geographical coverage of the sources, the time period searched, and the use, if any, of automated methods and tools.

Data Synthesis

We used narrative and tabulated methods for summarizing and synthesizing data collected during our data extraction stage. Whenever possible, charts have been used to visually present information relevant to the objectives of this study. We used the National Institute for Health and Care Excellence (NICE) disease classification system [14] to map the 96 studies (75%) [15-110] that reported specific disease indications or population groups. A total of 178 separate topic references throughout the 96 papers were extracted and mapped against the NICE classification. Despite some second-level disease indications (eg, COVID-19) that can be mapped to multiple NICE first-level disease classifications, each mention of disease was only mapped once to avoid inflation of results.

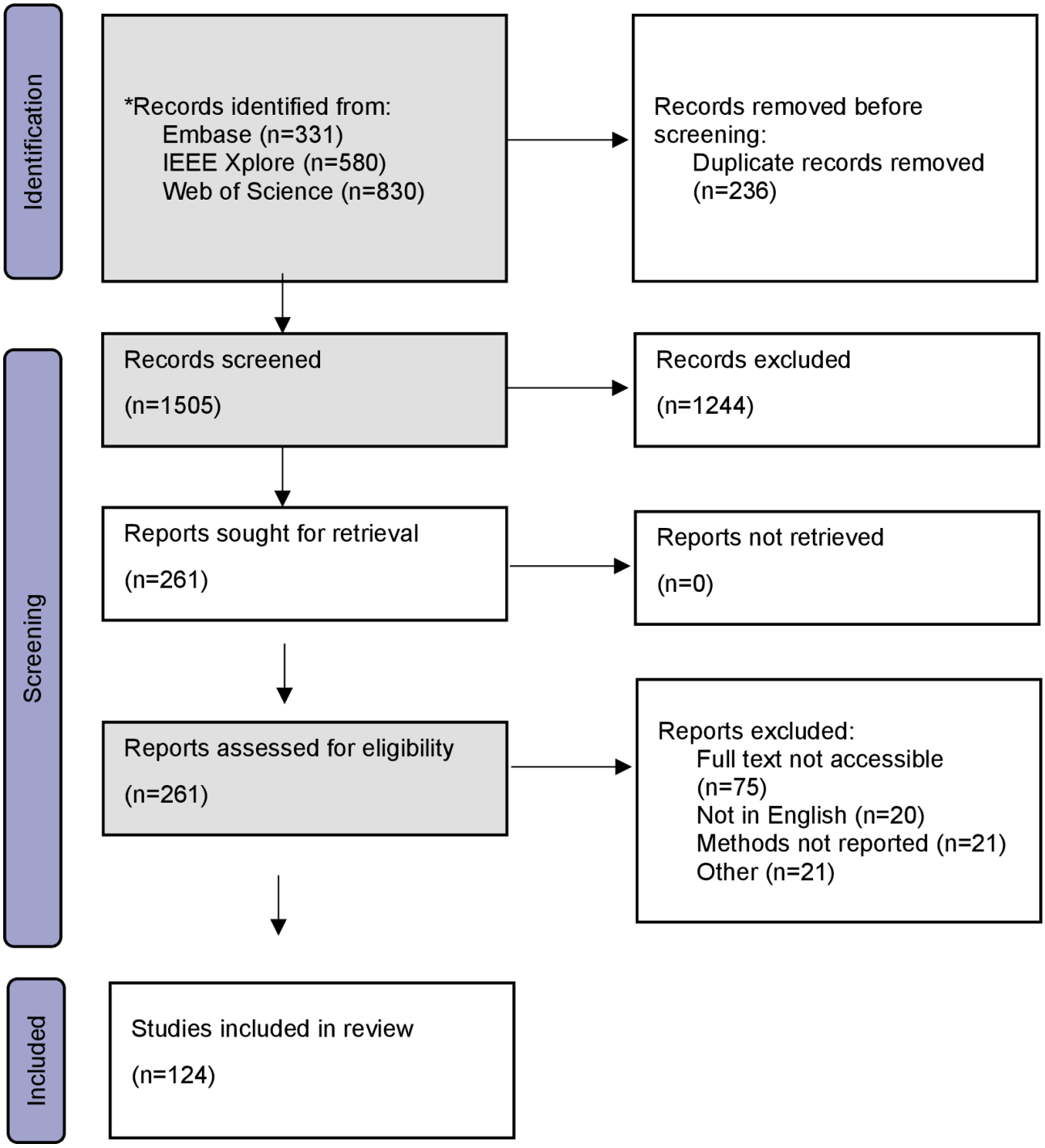
Results

Overview

Our searches identified a total of 1741 records. After deduplication, 1505 remained. We sought the full text of 261 records, of which only 124 were considered for inclusion in this review. Figure 1 presents a PRISMA flowchart of the retrieval and selection process for this review.

XSL•FO
RenderX

Figure 1. PRISMA flowchart for study selection via searching databases. Gray shading denotes rapid methods, for example, reduced number of databases searched, single screening, and single data extraction. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



The selection criteria focused on relevance and the potential to contribute to the review's objectives. This rigorous process ensured the reliability and comprehensiveness of the findings presented.

Study Characteristics

From the 124 studies included, 67 (54%) [15-61,111-130] were original research papers, 54 (43.5%) [62-107,131-138] corresponded to reviews, 2 were conference papers [108,109], and 1 short presentation [110]. The majority of sources considered patent databases only (85, 68%) [redacted] with the rest of the included studies using a combination of different sources. These 124 studies were published in 94 different journals. *Nature Biotechnology* was the most frequent journal in our dataset (n=9 studies) [15,23,24,52-54,123-125]. A table outlining the characteristics of the included studies is presented in Multimedia Appendix 2 [15-138].

Research Objectives of Included Studies

The objectives of each paper were extracted verbatim directly from the papers. In order to analyze these, and in line with recommendations on how to analyze and present data from scoping reviews [139], 3 reviewers (SG, EP, and OE) grouped them by common topics. This produced a final list of 9 topics or reasons (Table 1). These were presented and agreed with the

XSL•FO
RenderX

rest of the review team. The most common reason reported for the study of patents (62 studies, 50%) [5,7,9,14,17,21,23,27,28,31,33,36,38,39,41,43,44,46,48,50,52,53,55,58,60,62,63,65,66,68,69,72,73,75,77,78,80,82,84,85,87,88,90,91,93,94,96,97,99,101,103,104,106,107,109,111,112,114,116,117,119,120,122,123,125,126,128,129,131,132,134,136,137] was to investigate trends within specific fields, followed by those aiming to provide recommendations for future research, policy, and strategy development (25 studies, 20%) [20,22,23,25,26,32,33,35,44,121,58,62,127,69,70,76,80,84-86,89,91,92,100,135]. Two other frequent objectives were: the study of specific patented technologies (13 studies, 10%) [18,29,112,34,115,38,48,57,64,71,75,106,109] and the identification of emerging technologies (13 studies, 10%) [18,29,112,34,38,115,48,57,64,71,75,106,109]. Less frequent objectives included the development of methodological frameworks for patent analysis (4 studies) [56,88,117,122], analysis of global patent trends (3 studies) [50,96,137], cross-sectoral applications of emerging technologies (1 study) [55], summarize recent research in specific fields (1 study) [138], analyze patent application pipeline (1 study) [49], and develop patent research methods (1 study) [118].

Table . Reasons for conducting a patent research study by themes.

Patent analysis objective	Number of studies
Identify and analyze patent trends	62
Provide future directions	25
Identify emerging technologies	13
Analyze patented technologies	13
Methodological framework	4
Compare global patent trends	3
Cross-sectoral applications	1
Summarize recent research	1
Analyze patent application pipeline	1
Develop patent research methods	1

Sources of Patent Data

Eighty-five studies [5,7,9,14,17,21,23,27,28,31,33,36,38,39,41,43,44,46,48,50,52,53,55,58,60,62,63,65,66,68,69,72,73,75,77,78,80,82,84,85,87,88,90,91,93,94,96,97,99,101,103,104,106,107,109,111,112,114,116,117,119,120,122,123,125,126,128,129,131,132,134,136,137] used patents alone, whereas 39 studies [1,2,4,8,10,11,13,15,16,18,19,22,24,26,30,34,35,40,42,45,47,49,51,54,56,59,61,64,67,70,71,74,76,79,81,83,86,89,92,95,98,100,102,105,108,110,113,115,118,121,124,127,130,133,135,138] used various sources that included gray literature, research papers, and clinical trials. Two studies [89,96] did not report which patent database had been used. More than 27% (34/124) of papers described the use of more than 1 patent database. As such, across the 124 papers, there were 47 databases listed. Table 2 presents the distribution of the top 10 named databases across all papers.

Table . Top 10 patent databases used in the included studies.

Patent database	Number of papers
Derwent Innovation	30
USPTO[a]	24
Espacenet	20
The Lens	20
PatentScope	13
WIPO[b]	13
EPO[c]	10
Google Patents	10
Orbit	6
PatSeer	5

[a]USPTO: United States Patent and Trademark Office.

[b]WIPO: World Intellectual Property Office.

[c]EPO: European Patent Office.

XSL•FO
RenderX

Time Horizons of Patent Scans

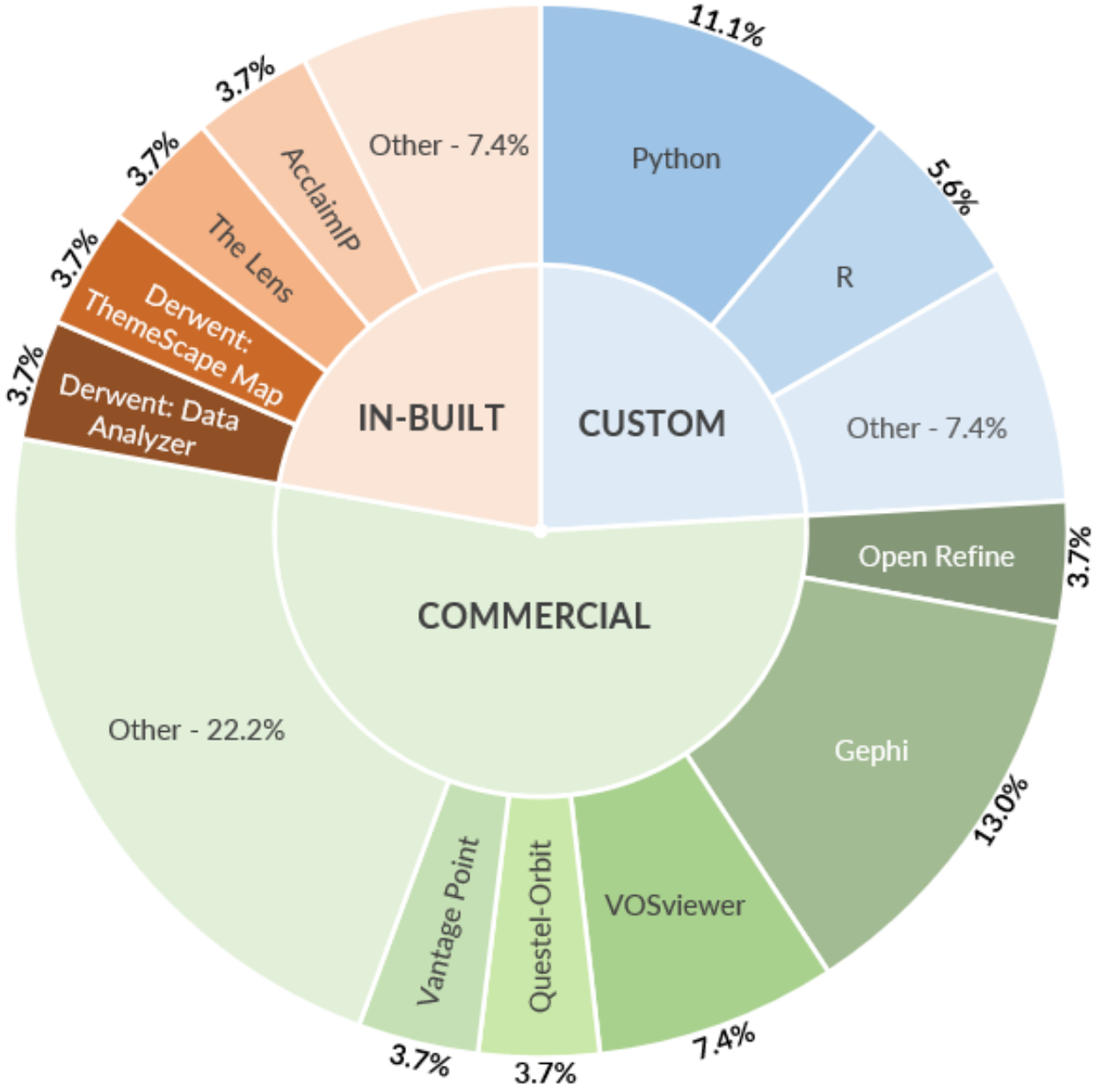
Of the 124 included papers, 44 (35.5%) papers [redacted] did not report the time period searched during patent analysis or gave incomplete timelines with only an end date leaving 80 papers [redacted] with completed data on time horizons.

The mean time length covered in those 80 papers was 24.6 years (SD 2.37) [redacted] 7 papers [21,29,41,46,91,106,131] searched a time period of over 50 years with the broadest time period searched being from 1900 to 2019 [140]. Around 19% (15/78) of papers with a reported time horizon searched a period between 3 months and 10 years. Up to 15% (12/78) of papers searched a 20-year period. It was common for authors to backdate searches to the start of the year 2000, although the rationale was not provided.

Methods and Automated Approaches Used

Of the 124 included papers, 43 (34.6%) studies [15-35,111-117,62-72,131,132,108,109] reported the use of automated or semiautomated approaches during patent analysis. In these 41 papers, there were 43 unique tools that were used for patent analysis. Figure 2 presents all tools categorized into whether the tool was available as an in-built analysis module by the patent database searched (IN-BUILT), a custom patent analysis script (CUSTOM), or a commercially offered tool (COMMERCIAL).

Figure 2. Reported tools used for automated or semiautomated analysis.



Some studies mentioned more than 1 tool, a full account of each tool used and their frequency distribution across the 41 studies is available in Table 3. These tools were used mainly in 4 different methods: data mining (understood as the statistical

XSL•FO
RenderX

technique of processing raw data in a structured form), text mining (the part of data mining which involves processing of text from documents) [141], data transformation (the process of converting and cleaning raw data from one data source to meet the requirements of its new location) [142], and network visualization (understood as the graphical representations of network devices, network metrics, and data flows) [143]. Network visualization was by far the most popular method, used in 50% of the tools described. The most popular tool for network visualization was the freely available software Gephi (Gephi Consortium), used in 7 separate studies [21-26,113]. Table 3 presents a breakdown of these tools classified by their type and the number of studies that used them.

XSL•FO
RenderX

Table . Full account of all named tools used for automated or semiautomated approaches for patent analysis from 41 papers that reported their use.

Tool	Number of studies used in	Tool type	Method	Reference
R – tidyverse	2	CUSTOM	Network visualization	[15,16]
R – patentr	1	CUSTOM	Network visualization	[116]
R – cowplot	1	CUSTOM	Network visualization	[16]
R – gridExtra	1	CUSTOM	Network visualization	[16]
R – lubridate	1	CUSTOM	Data transformation	[16]
R – magrittr	1	CUSTOM	Data transformation	[16]
R – linkcomm	1	CUSTOM	Network visualization	[16]
R – trend	1	CUSTOM	Data transformation	[16]
Python – pandas	3	CUSTOM	Data transformation	[16,17,111]
Python – patentpy	2	CUSTOM	Data transformation	[16,116]
Python – seaborn.clustermap	1	CUSTOM	Network visualization	[62]
Python-based Patent Enrichment Tool (PEMT)	1	CUSTOM	Data mining/data transformation	[62]
Python – gensim	1	CUSTOM	Data transformation	[112]
Python – matplotlib	1	CUSTOM	Network visualization	[111]
Latent Dirichlet Allocation algorithm	1	CUSTOM	Text mining	[18]
Word2Vec model	1	CUSTOM	Network visualization	[18]
Term Frequency-Inverse Document Frequency (TF-IDF) model	1	CUSTOM	Text mining	[19]
Generative Topographic Modeling	1	CUSTOM	Network visualization/data transformation	[20]
Gephi	7	COMMERCIAL	Network visualization	[21-26,113]
VOSviewer	4	COMMERCIAL	Network visualization	[27,64,108,131]
Open Refine	2	COMMERCIAL	Data transformation	[64,113]
Questel-Orbit	2	COMMERCIAL	Data mining	[28,132]
Vantage Point	2	COMMERCIAL	Text mining	[29,30]
UCINET 6	1	COMMERCIAL	Network visualization	[111]
Tableau	1	COMMERCIAL	Network visualization	[21]
Patentics	1	COMMERCIAL	Network visualization	[65]
Pajek	1	COMMERCIAL	Network visualization	[22]
Orange-Data Mining	1	COMMERCIAL	Text mining	[66]
Cytoscape	1	COMMERCIAL	Network visualization	[26]
Harzing's Publish or Perish	1	COMMERCIAL	Data mining	[67]
ITG Insight	1	COMMERCIAL	Text mining/network visualization	[114]
OriginPro	1	COMMERCIAL	Network visualization	[131]
Patsnap	1	COMMERCIAL	Network visualization	[31]
PatentInspiration	1	COMMERCIAL	Network visualization	[68]
NoteExpress	1	COMMERCIAL	Data transformation	[24]
Derwent: ThemeScape Map	2	IN-BUILT	Network visualization	[69,70]
Derwent: Data Analyzer	2	IN-BUILT	Data transformation	[69,108]
Lens	2	IN-BUILT	Data mining	[63,71]

XSL•FO
RenderX

Tool	Number of studies used in	Tool type	Method	Reference
AcclaimIP	1	IN-BUILT	Data mining	[32]
PatentScope	1	IN-BUILT	Text mining	[33]
Derwent: Smart Search	1	IN-BUILT	Data transformation	[115]
Derwent Innovation	1	IN-BUILT	Text mining	[34]
EP full-text search	1	IN-BUILT	Data mining	[72]

NICE Disease Classification Mapping

Of the 124 papers, 28 (22.6%) [111-116,117-122,123-129,130-132,133-138] did not sufficiently refer to any level of NICE classification. Between the remaining 96 papers [15-110], there were 177 specific mentions of NICE guidance categories, which are detailed below (Table 4). Overall, "Cancer" was the most prevalent NICE topic with 19.4% (19/96) of the 124 included studies reporting cancer patents as either the primary or secondary finding. Of the 16.1% of papers [19,26,28,30,31,33,44,76,84,96,100,104] that reported "Respiratory conditions," 60% (9/15)19,26,30,44,76,84,96,100,104 of these were specific to "COVID-19" or other coronaviruses.

Table . Specific mentions of NICE[a] disease and other guidance topics as a percentage of 178 NICE guidance topics referenced in the results of 94 papers that reported specific NICE topics in their results.

NICE guidance (first level)	Number of mentions	Percentage of papers
Cancer	24	19.4
Infections	20	16.1
Respiratory conditions	19	15.3
Diabetes and other endocrine, nutritional, and metabolic conditions	11	8.9
Injuries, accidents, and wounds	11	8.9
Neurological conditions	11	8.9
Blood and immune system conditions	10	8.1
Cardiovascular conditions	9	7.3
Oral and dental health	9	7.3
Skin conditions	8	6.5
Health and social care delivery	8	6.5
Digestive tract conditions	6	4.8
Mental health, behavioral, and neurodevelopmental disorders	6	4.8
Eye conditions	5	4
Musculoskeletal conditions	5	4
Urological conditions	4	3.2
Chronic and neuropathic pain	2	1.6
Liver conditions	2	1.6
Sleep and sleep conditions	2	1.6
Ear, nose, and throat conditions	1	0.8
Fertility, pregnancy, and childbirth	1	0.8
Lifestyle and well-being	1	0.8
Not reported	28	22.6

[a]NICE: National Institute for Health and Care Excellence.

XSL•FO
RenderX

Discussion

Principal Findings

Our rapid scoping review aimed to explore the recent literature on patent research studies to identify patent sources, tools, and methods used to understand how patent analyses may support horizon scanning methods in the detection of early innovations. To our knowledge, this is the first rapid scoping review focused on methodological approaches for patent research.

This review has identified several key findings that may be used to inform horizon scanning methods. First, our review has revealed an array of methodological approaches, tools, and reasons why patents are being used to discover signs of innovation. We identified 47 different sources for patent data retrieval with some studies using more than one source per study; for example, clinical trials and published or unpublished literature. The compilation of these sources in this review provides an up-to-date published list of patent data sources and insights into how researchers triangulate patent data with information on how, or whether, the innovation is being used in clinical studies and emerging research. This is an important finding for horizon scanning projects in general, and those undertaken by the IO, with particular value for the prioritization of scalable innovation. Patents alone may not be enough to assess the scalability of innovations in health care services such as the UK NHS. Combining patent data with clinical trials, publications, and clinical development analysis offers a more meaningful approach to evaluating technology adoption potential.

Second, this study showed that patent data are mostly used in trend analysis, perhaps due to the accessibility to retrospective global patent data. This finding aligns with what we already know about methods used for scanning emerging technologies and predicting innovations [3,144]. However, this review revealed a less common but notable application of using patent analyses to guide future policy, research, and strategy, similar to the "bottom-up" approach of the European Commission's Joint Research Centre for detecting weak signals [145]. Most importantly, this approach can also reveal research gaps, thereby aiding in directing future funding and policies to address these gaps, thus contributing to eliminating health inequities and disparities. These findings will contribute to consolidating the IO's methodological application of patent trend analysis studies to underpin "the push" or "bottom-up approach" to informing UK national future policy, research, or strategy on emerging innovation trends or identified innovation gaps.

Third, from our mapping of disease areas, not surprisingly, cancer, respiratory conditions, and infections were the top 3 topics addressed by the included studies. Considering that this rapid scoping review limited the included studies to only those published between 2020 and 2023, it was expected that COVID or, more broadly, respiratory conditions appeared in our dataset quite prominently. However, the focus on cancer technologies may respond to the growing burden of cancer worldwide and its impact on underserved populations [146]. Likewise, the emphasis on infection topics may respond to the antimicrobial resistance crisis that we are currently facing as well as the challenges posed by the growth of communicable diseases worldwide that continue to be the cause of substantial morbidity and mortality [147]. The alignment with real-world health challenges highlights innovators' responses and reinforces the need for horizon scanning systems to adopt these methods for studying innovation in major health issues.

Fourth, the analysis showed that tools used for patent data analysis primarily focus on creating graphs and network visualizations, with examples including Gephi [148], VOSViewer (Leiden University) [149], and ThemeScape Map (Derwent Innovation, Clarivate) [150]. Data manipulation tools such as Pandas (Wes McKinney in Python) [151] also played a key role. In addition, Python and R were the most common programming languages used to develop custom tools. Findings from this review highlighted that data transformation and visualization are methods often used for detecting emerging trends. This finding is in line with previously published studies into patent landscaping in the field of life sciences innovation [152]. Since there are no established methodological frameworks for analyzing patent data in horizon scanning, based on these results, it would be reasonable to assume that whenever tools are used for patent data analysis, some level of data transformation (or curation) is needed and therefore this step and the tools needed to undertake it should be part of the repertoire of methods for any horizon scanning practitioner and, more specifically, for the IO's methods toolkit.

Finally, our review identified a wide range of technologies for which patent analyses were undertaken (Table 3, Multimedia Appendix 2). Due to the short time frame of our analysis, we could not establish a methodological link between analysis methods and technology types. A longer time frame would have yielded more historical data that could have allowed some trend analysis on patent study methods by technology type. Notwithstanding, our results suggest the broad applicability of patent studies to any type of technology.

Our study has some limitations. First, we imposed a short time frame for our bibliographic database searches, which may have meant that we missed some relevant methodological studies. However, this was a pragmatic decision justified by the rapid nature of this review. Given the aims and objectives of our review, which consisted of finding methods and sources for patent landscape analysis studies, we do not think that limiting the search to the last 3 years of publication has biased our methodological conclusions; if anything, it has provided the most up-to-date view on tools and methods. Second, a longer time period would have provided more data allowing us to study the link between technology types and methods providing a more holistic view of patent analysis methods. Third, we used single screening to select studies, so we did not assess interrater agreement. While this practical choice may have led to some relevant studies being missed, we believe it would not have significantly affected our overall conclusions. Finally, our review was not able to ascertain if deduplication of data, a critical step to prevent data inflation and ensure accurate analyses of innovation clusters, was addressed by the included studies [152].

XSL•FO
RenderX

Conclusions

Our findings highlight a broad range of methodological approaches and tools which may suggest that in the health care technology context, the continuous search for signs of innovation is driving methods and tools development. Our study has shed light on the usefulness and purpose of patent research studies in the context of health care technology innovation, but it has not solved some methodological aspects such as deduplication processes and possible links between technology types and methods. Notwithstanding, when conducting such studies, particular attention should be given to the processes of data manipulation, transformation, and deduplication—especially when integrating multiple data sources—and these methods should be transparently documented and reported in the final output to ensure reproducibility and data quality. Implementing existing guidelines for reporting items for patent landscapes is also strongly recommended [2,153].

The IO aims to provide a systematic approach to horizon scanning across the 3 horizons of health care technology development: emerging, transitional, and imminent. The findings from this study will support the practice of horizon scanning in the emerging technology horizon. In summary, we expect that the findings from this rapid scoping review would be of value to a wide range of horizon scanning practitioners as well as those planning to undertake patent research studies in the field of health care innovation.

Acknowledgments

This study would not have been possible without the support of the Innovation Observatory. We thank Oleta Williams for the Information Specialist support provided for scoping and running the bibliographic databases searches for this rapid scoping review. This project is funded by the National Institute for Health and Care Research (NIHR) (HSRIC-2016 - 10009/Innovation Observatory).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SGGM and AM contributed to conceptualization, methodology, supervision, and project administration. SGGM, ER, OE, EG, JE, and AM were involved in data curation, formal analysis, validation, and visualization. All coauthors participated in writing – original draft, writing – review and editing, and approval of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategies.
[DOCX File, 33 KB - i-jmr_v14i1e70323_app1.docx]

Multimedia Appendix 2
Characteristics of included studies.
[DOCX File, 284 KB - i-jmr_v14i1e70323_app2.docx]

Checklist 1
PRISMA-ScR checklist.
[PDF File, 357 KB - i-jmr_v14i1e70323_app3.pdf]

References

1. Frequently asked questions: patents. World Intellectual Property Office (WIPO). URL: https://www.wipo.int/patents/en/faq_patents.html [accessed 2025-07-15]
2. Trippe A. Guidelines for preparing patent landscape reports. World Intellectual Property Office. 2015. URL: https://www.wipo.int/edocs/pubdocs/en/wipo_pub_946.pdf [accessed 2025-08-03]
3. Garcia Gonzalez-Moral S, Beyer FR, Oyewole AO, Richmond C, Wainwright L, Craig D. Looking at the fringes of MedTech innovation: a mapping review of horizon scanning and foresight methods. BMJ Open 2023 Sep 14;13(9):e073730. [doi: 10.1136/bmjopen-2023-073730] [Medline: 37709340]
4. Khan S, Fairbairn R, Potter R, Hussain A, Inskip A, Norman G. Horizon scanning methods for identification of new and repurposed medicines for stakeholders in the United Kingdom. Futures & Foresight Sci 2025 Apr;7(1):e210. [doi: 10.1002/ffo2.210]

XSL•FO
RenderX

5. Doos L, Packer C, Ward D, Simpson S, Stevens A. Past speculations of the future: a review of the methods used for forecasting emerging health technologies. BMJ Open 2016 Mar 10;6(3):e010479. [doi: 10.1136/bmjopen-2015-010479] [Medline: 26966060]
6. van Rijn T, Timmis JK. Patent landscape analysis-contributing to the identification of technology trends and informing research and innovation funding policy. Microb Biotechnol 2023 Apr;16(4):683-696. [doi: 10.1111/1751-7915.14201] [Medline: 36691841]
7. Garcia Gonzalez-Moral S, Pennock E, Ewedairo O, Green E, Elgey J, Mkwashi A. Healthcare technology innovation identification: a rapid scoping review of patent research studies to inform horizon scanning methods. Interact J Med Res. 2024 Jun 18. [doi: 10.2196/70323]
8. Garritty C, Hamel C, Trivella M, et al. Updated recommendations for the Cochrane Rapid Review Methods guidance for rapid reviews of effectiveness. BMJ 2024 Feb 6;384:e076335. [doi: 10.1136/bmj-2023-076335] [Medline: 38320771]
9. Campbell F. How to do a rapid scoping review. Cochrane. 2024. URL: https://training.cochrane.org/resource/how-to-do-a-rapid-scoping-review [accessed 2025-07-15]
10. Campbell F, Sutton A, Pollock D, et al. Rapid reviews methods series: guidance on rapid scoping, mapping and evidence and gap map ('Big Picture Reviews'). BMJ Evid Based Med 2025 Apr 21:bmjebm-2023-112389. [doi: 10.1136/bmjebm-2023-112389] [Medline: 39904600]
11. Higgins JPT, Green S. Guide to the contents of a cochrane protocol and review. In: Cochrane Handbook for Systematic Reviews of Interventions 2008:51-79. [doi: 10.1002/9780470712184]
12. 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]
13. 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]
14. NICE guidance. Conditions and diseases. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/conditions-and-diseases [accessed 2025-07-15]
15. Mohajel N, Arashkia A. Ebola as a case study for the patent landscape of medical countermeasures for emerging infectious diseases. Nat Biotechnol 2021 Jul;39(7):799-807. [doi: 10.1038/s41587-021-00970-z] [Medline: 34257454]
16. Wadhwa RR, McElderry BM, Yu J, et al. Temporal trends in the United States patent landscape: innovation in cardiology across industry and academia. Cardiol Res 2023 Oct;14(5):334-341. [doi: 10.14740/cr1417] [Medline: 37936628]
17. Rincón-López J, Almanza-Arjona YC, Riascos AP, Rojas-Aguirre Y. When cyclodextrins met data science: unveiling their pharmaceutical applications through network science and text-mining. Pharmaceutics 2021 Aug 19;13(8):1297. [doi: 10.3390/pharmaceutics13081297] [Medline: 34452258]
18. Shin HJ, Lee S. Identifying areas of technology commercialization in the biomedical sector: an integrated analysis of patents and publications. Presented at: 2022 Portland International Conference on Management of Engineering and Technology (PICMET); Aug 7-11, 2022; Portland, OR. [doi: 10.23919/PICMET53225.2022.9882635]
19. Durmuşoğlu A, Durmuşoğlu ZDU. Remembering medical ventilators and masks in the days of COVID-19: patenting in the last decade in respiratory technologies. IEEE Trans Eng Manage 2022;71:1359-1373. [doi: 10.1109/TEM.2022.3151636]
20. Hwang J, Kim KH, Hwang JG, Jun S, Yu J, Lee C. Technological opportunity analysis: assistive technology for blind and visually impaired people. Sustainability 2020;12(20):8689. [doi: 10.3390/su12208689]
21. Xin Y, Man W, Yi Z. The development trend of artificial intelligence in medical: a patentometric analysis. Artif Intell Life Sci 2021 Dec;1:100006. [doi: 10.1016/j.ailsci.2021.100006]
22. Cai Y, Zhang X, Zhang K, et al. The global patent landscape of emerging infectious disease monkeypox. BMC Infect Dis 2024 Apr 15;24(1):403. [doi: 10.1186/s12879-024-09252-w] [Medline: 38622539]
23. Li Q, Xu J, Sun Q, Zhang Z, Hu Y, Yao H. The global patent landscape of HER2-targeted biologics. Nat Biotechnol 2023 Jun;41(6):756-764. [doi: 10.1038/s41587-023-01814-8] [Medline: 37316732]
24. Lyu M, Chen J, Peng Y, et al. The global patent landscape of mRNA for diagnosis and therapy. Nat Biotechnol 2023 Sep;41(9):1193-1199. [doi: 10.1038/s41587-023-01925-2] [Medline: 37699990]
25. Ma J, Pan YH, Su CY. Organization-oriented technology opportunities analysis based on predicting patent networks: a case of Alzheimer's disease. Scientometrics 2022 Sep;127(9):5497-5517. [doi: 10.1007/s11192-021-04219-z]
26. Liu K, Zhang X, Hu Y, et al. What, where, when and how of COVID-19 patents landscape: a bibliometrics review. Front Med (Lausanne) 2022;9:925369. [doi: 10.3389/fmed.2022.925369] [Medline: 35847804]
27. Wang Q, Zhu Q, Li N. A scientometric analysis and visualization of scientific research and technology innovation in needle-free insulin injection from 1974 to 2022. Clin Ther 2023 Sep;45(9):881-888. [doi: 10.1016/j.clinthera.2023.06.025] [Medline: 37516566]
28. Machuca-Martinez F, Amado RC, Gutierrez O. Coronaviruses: a patent dataset report for research and development (R&D) analysis. Data Brief 2020 Apr;30:105551. [doi: 10.1016/j.dib.2020.105551] [Medline: 32337328]
29. Klongthong W, Muangsin V, Gowanit C, Muangsin N. A patent analysis to identify emergent topics and convergence fields: a case study of chitosan. Sustainability 2021;13(16):9077. [doi: 10.3390/su13169077]

XSL•FO
RenderX

30. Devarapalli P, Kumari P, Soni S, Mishra V, Yadav S. Patent intelligence of RNA viruses: implications for combating emerging and re-emerging RNA virus based infectious diseases. Int J Biol Macromol 2022 Oct 31;219:1208-1215. [doi: 10.1016/j.ijbiomac.2022.08.169] [Medline: 36058387]
31. Zhang HL, Li YX, Zhou AF, Li Y. New frontier in antiviral drugs for disorders of the respiratory system. Recent Adv Antiinfect Drug Discov 2022;17(1):2-12. [doi: 10.2174/1574891X16666220416164740] [Medline: 35430978]
32. Altuntas S, Erdogan Z, Dereli T. A clustering-based approach for the evaluation of candidate emerging technologies. Scientometrics 2020 Aug;124(2):1157-1177. [doi: 10.1007/s11192-020-03535-0]
33. Yeh TF, Lin C, Sung HC. A review of technological developments in lipid nanoparticle application for mRNA vaccination. Hum Vaccin Immunother 2023 Aug 1;19(2):2256040. [doi: 10.1080/21645515.2023.2256040] [Medline: 37705354]
34. Abdi S, Kitsara I, Hawley MS, de Witte LP. Emerging technologies and their potential for generating new assistive technologies. Assist Technol 2021 Dec 1;33(sup1):17-26. [doi: 10.1080/10400435.2021.1945704] [Medline: 34951831]
35. Jeon E, Yoon N, Sohn SY. Exploring new digital therapeutics technologies for psychiatric disorders using BERTopic and PatentSBERTa. Technol Forecast Soc Change 2023 Jan;186:122130. [doi: 10.1016/j.techfore.2022.122130]
36. Oda T, Oda C. An analysis of the key drivers of the Japanese digital therapeutics patents: a cross‐sectional study. Health Sci Rep 2023 Jun;6(6):e1268 [FREE Full text] [doi: 10.1002/hsr2.1268] [Medline: 37275666]
37. Valadas LAR, Oliveira Filho RD, Lotif MAL, et al. Development and innovation on dental products in Argentina: a technological prospecting based on patents.. Recent Pat Biotechnol 2021;15(1):3-11. [doi: 10.2174/1872208314999201210222828] [Medline: 33305714]
38. Juiz PJL, Ferreira LTB, Pires EA, Villarreal CF. Patent mining on the use of antioxidant phytochemicals in the technological development for the prevention and treatment of periodontitis. Antioxidants (Basel) 2024 May 3;13(5):566. [doi: 10.3390/antiox13050566] [Medline: 38790671]
39. Chartoumpekis DV, Fu CY, Ziros PG, Sykiotis GP. Patent review (2017–2020) of the Keap1/Nrf2 pathway using PatSeer Pro: focus on autoimmune diseases. Antioxidants (Basel) 2020 Nov 17;9(11):1138. [doi: 10.3390/antiox9111138] [Medline: 33212784]
40. Speziali MG. Cellulose technologies applied to biomedical purposes from the patentometric point of view. Cellulose 2020 Nov;27(17):10095-10117. [doi: 10.1007/s10570-020-03477-z]
41. Cho YD, Kim WJ, Ryoo HM, Ku Y. Patent landscape report on dental implants: a technical analysis. Clin Implant Dent Relat Res 2021 Dec;23(6):857-863 [FREE Full text] [doi: 10.1111/cid.13048] [Medline: 34669245]
42. Sharma R, Sharma R, Singla RK. Drug discovery, diagnostic, and therapeutic trends on Mpox: a patent landscape. Curr Res Biotechnol 2024;7:100173. [doi: 10.1016/j.crbiot.2023.100173]
43. Tiwari AK, Jaspal D, Deshmukh S, et al. Cheminformatics: a patentometric analysis. Fr Ukr J Chem 2022;10(1):13-29. [doi: 10.17721/fujcV10I1P13-29]
44. Bacigalupo ML, Pignataro MF, Scopel CT, Kondratyuk S, Mellouk O, Chaves GC. Unveiling patenting strategies of therapeutics and vaccines: evergreening in the context of COVID-19 pandemic. Front Med (Lausanne) 2023;10:1287542. [doi: 10.3389/fmed.2023.1287542] [Medline: 38126073]
45. Li M, Ren J, Si X, et al. The global mRNA vaccine patent landscape. Hum Vaccin Immunother 2022 Nov 30;18(6):2095837. [doi: 10.1080/21645515.2022.2095837] [Medline: 35797353]
46. Xiong YH, Xu XN, Zheng B. Patented technologies for schistosomiasis control and prevention filed by Chinese applicants. Infect Dis Poverty 2021 Jun 12;10(1):84. [doi: 10.1186/s40249-021-00869-6] [Medline: 34118989]
47. Francis N, Ong A, Suhaimi H, Abas PE. Patent landscape review on ankle sprain prevention method: technology updates. Inventions 2023;8(2):53. [doi: 10.3390/inventions8020053]
48. Wang YH. Exploring Technology-Driven Technology Roadmaps (TRM) for wearable biosensors in healthcare. IRBM 2024 Jun;45(3):100835. [doi: 10.1016/j.irbm.2024.100835]
49. Ge J. Analysis of patent development status of lipid nanoparticle delivery system for mRNA vaccines. J Chin Pharm Sci 2023 Feb 28;32(2):112. [doi: 10.5246/jcps.2023.02.009]
50. Xu C, Zhu D. The development of marine drugs: a research based on patent analysis. J Coast Res 2020;110(sp1):271-276. [doi: 10.2112/JCR-SI110-064.1]
51. Riondato M, Rigamonti D, Martini P, et al. Oldie but goodie: is Technetium-99m still a treasure trove of innovation for medicine? A patents analysis (2000-2022). J Med Chem 2023 Apr 13;66(7):4532-4547. [doi: 10.1021/acs.jmedchem.3c00148] [Medline: 37010457]
52. Lyu L, Feng Y, Chen X, Hu Y. The global chimeric antigen receptor T (CAR-T) cell therapy patent landscape. Nat Biotechnol 2020 Dec;38(12):1387-1394. [doi: 10.1038/s41587-020-00749-8] [Medline: 33273733]
53. Braga L, Lopes R, Alves L, Mota F. The global patent landscape of artificial intelligence applications for cancer. Nat Biotechnol 2023 Dec;41(12):1679-1687. [doi: 10.1038/s41587-023-02051-9] [Medline: 38082076]
54. Greenberg A, Cohen A, Grewal M. Patent landscape of brain-machine interface technology. Nat Biotechnol 2021 Oct;39(10):1194-1199. [doi: 10.1038/s41587-021-01071-7] [Medline: 34621062]
55. Lee S, Hwang J, Cho E. Comparing technology convergence of artificial intelligence on the industrial sectors: two-way approaches on network analysis and clustering analysis. Scientometrics 2022 Jan;127(1):407-452. [doi: 10.1007/s11192-021-04170-z]

XSL•FO
RenderX

56. Hani U, Chen JW, Holland C, et al. Patent bibliometrics in spinal deformity: the first bibliometric analysis of spinal deformity's technological literature. Spine Deform 2024 Jan;12(1):25-33. [doi: 10.1007/s43390-023-00767-x] [Medline: 37845600]
57. Liu K, Yang J, Deng J, Fan X, Hu Y. Global patent landscape of benign prostatic hyperplasia drugs. Urology 2022 Aug;166:209-215. [doi: 10.1016/j.urology.2022.03.008] [Medline: 35314182]
58. Parihar K, Telang M, Ovhal A. A patent review on strategies for biological control of mosquito vector. World J Microbiol Biotechnol 2020 Nov 24;36(12):187. [doi: 10.1007/s11274-020-02960-w] [Medline: 33230723]
59. Borzova E, Cardeal G, Soperna S, Zhao J, Lepekhova A. 582 The patent landscape analysis of skin bioinks for 3D bioprinting. J Invest Dermatol 2022 Dec;142(12):S281. [doi: 10.1016/j.jid.2022.09.598]
60. Lee JH, Kim JS, Moon BC, Nam HH. Analysis of trends in patents on insect-derived medicinal materials for skin diseases. Integr Med Res 2020;9:100617. [doi: 10.1016/j.imr.2020.100617]
61. Litvinova O, Bilir A, Parvanov ED, et al. Patent landscape review of non-invasive medical sensors for continuous monitoring of blood pressure and their validation in critical care practice. Front Med (Lausanne) 2023;10:1138051. [doi: 10.3389/fmed.2023.1138051] [Medline: 37497278]
62. Gadiya Y, Gribbon P, Hofmann-Apitius M, Zaliani A. Pharmaceutical patent landscaping: a novel approach to understand patents from the drug discovery perspective. Artif Intell Life Sci 2023 Dec;3:100069. [doi: 10.1016/j.ailsci.2023.100069]
63. Litvinova O, Eitenberger M, Bilir A, et al. Patent analysis of digital sensors for continuous glucose monitoring. Front Public Health 2023;11:1205903. [doi: 10.3389/fpubh.2023.1205903] [Medline: 37621612]
64. Zagoya-López J, Zúñiga-Avilés LA, Vilchis-González AH, Ávila-Vilchis JC. Foot/ankle prostheses design approach based on scientometric and patentometric analyses. Appl Sci (Basel) 2021;11(12):5591. [doi: 10.3390/app11125591]
65. Jiang J, Sun J, Huang Z, et al. The state of the art and future trends of root canal files from the perspective of patent analysis: a study design. Biomed Eng Online 2022 Dec 24;21(1):90. [doi: 10.1186/s12938-022-01060-0] [Medline: 36566212]
66. Patel S, Patel MS, Patel AD, et al. Probiotic formulations: a patent landscaping using the text mining approach. Curr Microbiol 2022 Apr 9;79(5):152. [doi: 10.1007/s00284-022-02836-2] [Medline: 35397006]
67. Azman AA, Leow ATC, Noor NDM, Noor SAM, Latip W, Ali MSM. Worldwide trend discovery of structural and functional relationship of metallo-β-lactamase for structure-based drug design: a bibliometric evaluation and patent analysis. Int J Biol Macromol 2024 Jan;256(Pt 2):128230. [doi: 10.1016/j.ijbiomac.2023.128230] [Medline: 38013072]
68. Cañete R, Peralta E. Assistive technology to improve collaboration in children with ASD: state-of-the-art and future challenges in the smart products sector. Sensors (Basel) 2022 Oct 30;22(21):8321. [doi: 10.3390/s22218321] [Medline: 36366019]
69. Zhou W, Wang X. Human gene therapy: a patent analysis. Gene 2021 Nov 30;803:145889. [doi: 10.1016/j.gene.2021.145889] [Medline: 34371094]
70. Kim WJ, Cho YD, Ku Y, Ryoo HM. The worldwide patent landscape of dental implant technology. Biomater Res 2022 Oct 23;26(1):59. [doi: 10.1186/s40824-022-00307-0] [Medline: 36274171]
71. Bhatnagar P, Law JX, Ng SF. Delivery systems for platelet derived growth factors in wound healing: a review of recent developments and global patent landscape. J Drug Deliv Sci Technol 2022 May;71:103270. [doi: 10.1016/j.jddst.2022.103270]
72. Gkika DA, Vordos N, Maragakis M, et al. Patents of nanomaterials related with cancer treatment applications. J Nanopart Res 2020 Oct;22(10). [doi: 10.1007/s11051-020-05052-w]
73. Litvinova O, Klager E, Tzvetkov NT, et al. Digital pills with ingestible sensors: patent landscape analysis.. Pharmaceuticals (Basel) 2022 Aug 19;15(8):1025. [doi: 10.3390/ph15081025] [Medline: 36015173]
74. Maresova P, Krejcar O, Barakovic S, et al. Health–related ICT solutions of smart environments for elderly–systematic review. IEEE Access 2020;8:54574-54600. [doi: 10.1109/ACCESS.2020.2981315]
75. Wu NJW, Aquilina M, Qian BZ, et al. The application of nanotechnology for quantification of circulating tumour DNA in liquid biopsies: a systematic review. IEEE Rev Biomed Eng 2023;16:499-513. [doi: 10.1109/RBME.2022.3159389] [Medline: 35302938]
76. Imran M, Fatima W, Alzahrani AK, et al. Development of therapeutic and prophylactic zinc compositions for use against COVID-19: a glimpse of the trends, inventions, and patents. Nutrients 2022 Mar 14;14(6):1227. [doi: 10.3390/nu14061227] [Medline: 35334884]
77. Liu K, Zuo H, Li G, Yu H, Hu Y. Global research on artemisinin and its derivatives: perspectives from patents. Pharmacol Res 2020 Sep;159:105048. [doi: 10.1016/j.phrs.2020.105048] [Medline: 32590098]
78. Burgio V, Bei J, Rodriguez Reinoso M, et al. Mechanical stapling devices for soft tissue repair: a review of commercially available linear, linear cutting, and circular staplers. Appl Sci (Basel) 2024;14(6):2486. [doi: 10.3390/app14062486]
79. Singh M, Jassal R, Khetarpal P. Diagnostic and therapeutic approaches for endometriosis: a patent landscape. Arch Gynecol Obstet 2024 Mar;309(3):831-842. [doi: 10.1007/s00404-023-07151-0] [Medline: 37626175]
80. Imran M, Khan SA, Alshammari MK, et al. Discovery, development, inventions, and patent trends on mobocertinib succinate: the first-in-class oral treatment for NSCLC with EGFR exon 20 insertions. Biomedicines 2021 Dec 17;9(12):1938. [doi: 10.3390/biomedicines9121938] [Medline: 34944754]

XSL•FO
RenderX

81. Ailia MJ, Thakur N, Abdul-Ghafar J, Jung CK, Yim K, Chong Y. Current trend of artificial intelligence patents in digital pathology: a systematic evaluation of the patent landscape. Cancers (Basel) 2022 May 13;14(10):2400. [doi: 10.3390/cancers14102400] [Medline: 35626006]
82. Chuah LH, Loo HL, Goh CF, Fu JY, Ng SF. Chitosan-based drug delivery systems for skin atopic dermatitis: recent advancements and patent trends. Drug Deliv Transl Res 2023 May;13(5):1436-1455. [doi: 10.1007/s13346-023-01307-w] [Medline: 36808298]
83. Liu XX, Chen CY, Li L, et al. Bibliometric study of adaptogens in dermatology: pharmacophylogeny, phytochemistry, and pharmacological mechanisms.. Drug Des Devel Ther 2023;17:341-361. [doi: 10.2147/DDDT.S395256] [Medline: 36776447]
84. Zhou W, Chen D. Emerging patent landscape for gene therapy as a potential cure for COVID-19. Math Probl Eng 2021 Jan 8;2021:1-6. [doi: 10.1155/2021/6630419]
85. Yuan Y, Fu Q, Zhang Y, et al. CAR-based cell therapy: evaluation with bibliometrics and patent analysis. Hum Vaccin Immunother 2021 Nov 2;17(11):4374-4382. [doi: 10.1080/21645515.2021.1947100] [Medline: 34184975]
86. Picanco-Castro V, Gonçalves Pereira C, Swiech K, Ribeiro Malmegrim KC, Tadeu Covas D, Silveira Porto G. Emerging CAR T cell therapies: clinical landscape and patent technological routes. Hum Vaccin Immunother 2020 Jun 2;16(6):1424-1433. [doi: 10.1080/21645515.2019.1689744] [Medline: 31702480]
87. Asif M, Tiwana MI, Khan US, et al. Advancements, trends and future prospects of lower limb prosthesis. IEEE Access 2021;9:85956-85977. [doi: 10.1109/ACCESS.2021.3086807] [Medline: 36345377]
88. Culmone C, Yikilmaz FS, Trauzettel F, Breedveld P. Follow-the-leader mechanisms in medical devices: a review on scientific and patent literature. IEEE Rev Biomed Eng 2023;16:439-455. [doi: 10.1109/RBME.2021.3113395] [Medline: 34543205]
89. Shivakumar P, Gupta MS, Jayakumar R, Gowda DV. Prospection of chitosan and its derivatives in wound healing: proof of patent analysis (2010-2020). Int J Biol Macromol 2021 Aug 1;184:701-712. [doi: 10.1016/j.ijbiomac.2021.06.086] [Medline: 34157330]
90. Kurakula M, N NR. Prospection of recent chitosan biomedical trends: evidence from patent analysis (2009-2020). Int J Biol Macromol 2020 Dec 15;165(Pt B):1924-1938. [doi: 10.1016/j.ijbiomac.2020.10.043] [Medline: 33068625]
91. Liu K, Gu Z, Islam MS, et al. Global landscape of patents related to human coronaviruses. Int J Biol Sci 2021;17(6):1588-1599. [doi: 10.7150/ijbs.58807] [Medline: 33907523]
92. Lohita S, T RH, Rebello AP. Myocardial infraction: background, recent advances and interventions supported by clinical trial and patent landscape. Int J Pharm Sci Rev Res 2023;83(1):142-157. [doi: 10.47583/ijpsrr.2023.v83i01.024]
93. Queiroz A, Mendes IAC, de Godoy S, Velez Lapão L, Dias S. mHealth strategies related to HIV postexposure prophylaxis knowledge and access: systematic literature review, technology prospecting of patent databases, and systematic search on app stores. JMIR Mhealth Uhealth 2021 Feb 16;9(2):e23912. [doi: 10.2196/23912] [Medline: 33591289]
94. Kong X, Zuo H, Huang HD, et al. STING as an emerging therapeutic target for drug discovery: perspectives from the global patent landscape. J Adv Res 2023 Feb;44:119-133. [doi: 10.1016/j.jare.2022.05.006] [Medline: 35636721]
95. Singh P, Dhankhar J, Kapoor RK, et al. Ficus benghalensis—a comprehensive review on pharmacological research, nanotechnological applications, and patents. J Appl Pharm Sci 2023;13(10):59-82. [doi: 10.7324/JAPS.2023.134426]
96. Wei F, Zhou H, Gao G, Zheng Q. Analysis of trends in patent development for coronavirus detection, prevention, and treatment technologies in key countries. J Biosaf Biosecur 2022 Jun;4(1):23-32. [doi: 10.1016/j.jobb.2021.10.005] [Medline: 34746687]
97. Islam MM, Naveen NR, Anitha P, et al. The race to replace PDE5i: recent advances and interventions to treat or manage erectile dysfunction: evidence from patent landscape (2016-2021). J Clin Med 2022 May 31;11(11):3140. [doi: 10.3390/jcm11113140] [Medline: 35683526]
98. DasNandy A, Virge R, Hegde HV, Chattopadhyay D. A review of patent literature on the regulation of glucose metabolism by six phytocompounds in the management of diabetes mellitus and its complications. J Integr Med 2023 May;21(3):226-235. [doi: 10.1016/j.joim.2023.02.003] [Medline: 36932029]
99. Silva MB, Feitosa AO, Lima IGO, et al. Antarctic organisms as a source of antimicrobial compounds: a patent review. An Acad Bras Cienc 2022;94(suppl 1):e20210840. [doi: 10.1590/0001-3765202220210840] [Medline: 35384978]
100. Imran M, Kumar Arora M, Asdaq SMB, et al. Discovery, development, and patent trends on molnupiravir: a prospective oral treatment for COVID-19. Molecules 2021 Sep 24;26(19):5795. [doi: 10.3390/molecules26195795] [Medline: 34641339]
101. Raghu Kiran CVS, Gopinath C. Idiom of gastroretentive drug delivery systems: comprehensive view on innovation technologies, patents and clinical trails. Neuroquantology 2023;21(6):1237-1253. [doi: 10.48047/nq.2023.21.6.NQ23127]
102. Imran M, Asdaq SMB, Khan SA, et al. Innovations and patent trends in the development of USFDA approved protein kinase inhibitors in the last two decades. Pharmaceuticals (Basel) 2021 Jul 22;14(8):710. [doi: 10.3390/ph14080710] [Medline: 34451807]
103. Cammarano A, Dello Iacono S, Meglio C, Nicolais L. Advances in transdermal drug delivery systems: a bibliometric and patent analysis. Pharmaceutics 2023 Dec 12;15(12):2762. [doi: 10.3390/pharmaceutics15122762] [Medline: 38140102]
104. Mitsumori Y, Kato H, Kato A, Kamijo K. An analysis of COVID-19 related IPRS: should they be promoted, waived or pooled? Presented at: 2022 Portland International Conference on Management of Engineering and Technology (PICMET); Portland, OR p. 1-11. [doi: 10.23919/PICMET53225.2022.9882853]

XSL•FO
RenderX

105. Ali Hazis NU, Aneja N, Rajabalaya R, David SR. Systematic patent review of nanoparticles in drug delivery and cancer therapy in the last decade. Recent Adv Drug Deliv Formul 2021;15(1):59-74. [doi: 10.2174/1872211314666210521105534] [Medline: 34602031]
106. Russo Serafini M, Medeiros Savi F, Ren J, et al. The patenting and technological trends in hernia mesh implants. Tissue Eng Part B Rev 2021 Feb;27(1):48-73. [doi: 10.1089/ten.TEB.2019.0245] [Medline: 32403989]
107. Fluit R, Prinsen EC, Wang S, van der Kooij H. A comparison of control strategies in commercial and research knee prostheses. IEEE Trans Biomed Eng 2020 Jan;67(1):277-290. [doi: 10.1109/TBME.2019.2912466] [Medline: 31021749]
108. Zhang T, Chen J, Lu Y, Xu D, Yang X, Ouyang Z. The research activities and development trends of antineoplastics targeting PD-1/PD-L1 based on scientometrics and patentometrics. 2022 Presented at: 2022 International Conference on Computational Science and Computational Intelligence (CSCI); Dec 14-16, 2022; Las Vegas, NV. [doi: 10.1109/CSCI58124.2022.00300]
109. Verma R. A technical analysis of MIOT in sensitive aspect. Presented at: 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE); May 12-13, 2023; Greater Noida, India. [doi: 10.1109/ICACITE57410.2023.10182684]
110. Sertkaya A, McGeeney JD, Sullivan C, et al. Assessing the state of antibacterial drug discovery through patent analysis. Int J Antimicrob Agents 2024 Feb;63(2):107051. [doi: 10.1016/j.ijantimicag.2023.107051] [Medline: 38072169]
111. Chen TA, Tsai SC, Hu KK. Using big data analytics on health industry development: the empirical intellectual property analysis from stem cell therapy. Presented at: 2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS); May 28-30, 2021; Tainan, Taiwan. [doi: 10.1109/ECBIOS51820.2021.9510478]
112. Jeon D, Ahn JM, Kim J, Lee C. A doc2vec and local outlier factor approach to measuring the novelty of patents. Technol Forecast Soc Change 2022 Jan;174:121294. [doi: 10.1016/j.techfore.2021.121294]
113. Chen Y, Xiong SH, Li F, et al. Delivery of therapeutic small interfering RNA: the current patent-based landscape. Mol Ther Nucleic Acids 2022 Sep 13;29:150-161. [doi: 10.1016/j.omtn.2022.06.011] [Medline: 35847171]
114. Liu J, Wei JQ, Liu YQ. Technology forecasting based on topic analysis and social network analysis: a case study focusing on gene editing patents. J Sci Ind Res 2021 May;80(5):428-437. [doi: 10.56042/jsir.v80i05.40860]
115. Chang SH. The development trend and academic patent technology network of laser and optical technologies. Int J Innov Sci 2022 Mar 23;14(2):302-315. [doi: 10.1108/IJIS-02-2021-0044]
116. Wadhwa R, Yu J, Parikh A. eP277: Exploring the landscape of genetics patents in the United States from 2005 to 2020. Genet Med 2022 Mar;24(3):S174-S175. [doi: 10.1016/j.gim.2022.01.312]
117. Erzurumlu SS, Pachamanova D. Topic modeling and technology forecasting for assessing the commercial viability of healthcare innovations. Technol Forecast Soc Change 2020 Jul;156:120041. [doi: 10.1016/j.techfore.2020.120041]
118. Borge L, Wustmans M, Bröring S. Assessing interdisciplinary research within an emerging technology network: a novel approach based on patents in the field of bioplastics. IEEE Trans Eng Manage 2022;71:1452-1469. [doi: 10.1109/TEM.2022.3146199]
119. Robinson A, Islam N, Sengoku S, Miyazaki K. Examining the role of actors in an emerging technological system: the case of POC devices. Presented at: 2023 Portland International Conference on Management of Engineering and Technology (PICMET); Jul 23-27, 2023; Monterrey, Mexico. [doi: 10.23919/PICMET59654.2023.10216892]
120. Pasek JE, University of Wyoming, Laramie, WY, USA. Trends in bioengineering patents granted 2000 - 2019. Biomed Sci Instrum 2021 Apr 1;57(2):61-73. [doi: 10.34107/YHPN9422.0461]
121. Hernández-Melchor D, López-Bayghen E, Padilla-Viveros A. The patent landscape in the field of stem cell therapy: closing the gap between research and clinic. F1000Res 2023;11:997. [doi: 10.12688/f1000research.123799.4] [Medline: 38481536]
122. Mancilla-de-la-Cruz J, Rodriguez-Salvador M, Ruiz-Cantu L. The next pharmaceutical path: determining technology evolution in drug delivery products fabricated with additive manufacturing. Foresight and STI Governance 2020;14(3):55-70. [doi: 10.17323/2500-2597.2020.3.55.70]
123. Picanço-Castro V, Pereira CG, Covas DT, Porto GS, Athanassiadou A, Figueiredo ML. Emerging patent landscape for non-viral vectors used for gene therapy. Nat Biotechnol 2020 Feb;38(2):151-157. [doi: 10.1038/s41587-019-0402-x] [Medline: 32034383]
124. Aboy M, Liddell K, Liddicoat J, Crespo C, Jordan M. Mapping the European patent landscape for medical uses of known products. Nat Biotechnol 2021 Nov;39(11):1336-1343. [doi: 10.1038/s41587-021-01104-1] [Medline: 34754112]
125. Aboy M, Price WN, Raker S. Mapping the patent landscape of medical machine learning. Nat Biotechnol 2023 Apr;41(4):461-468. [doi: 10.1038/s41587-023-01735-6] [Medline: 37069385]
126. Frisio DG, Ventura V. Global innovation trends for plant-based vaccines production: a patent analysis. Plants (Basel) 2021 Nov 23;10(12):2558. [doi: 10.3390/plants10122558] [Medline: 34961028]
127. Danylenko Y, Nepokupna T. Analysis of scintillation materials for nuclear medicine on the basis of patent analytics. Sci Innov 2023;19(5):43-56. [doi: 10.15407/scine19.05.043]
128. Barragán-Ocaña A, Oliver-Espinoza R, Longar-Blanco MDP, Gómez-Viquez H. Technological development and patent analysis: the case of biopharmacy in the world and in Latin America. Tapuya 2022 Dec 31;5(1). [doi: 10.1080/25729861.2022.2111112]

XSL•FO
RenderX

129. Orozco Colonia BS, Vinícius de Melo Pereira G, Soccol CR. Omega-3 microbial oils from marine thraustochytrids as a sustainable and technological solution: a review and patent landscape. Trends Food Sci Technol 2020 May;99:244-256. [doi: 10.1016/j.tifs.2020.03.007]
130. Gadiya Y, Shetty S, Hofmann-Apitius M, Gribbon P, Zaliani A. Exploring SureChEMBL from a drug discovery perspective. Sci Data 2024 May 16;11(1):507. [doi: 10.1038/s41597-024-03371-4] [Medline: 38755219]
131. Almeida FLC, Castro MPJ, Travalia BM, et al. Trends in lipase immobilization: bibliometric review and patent analysis. Process Biochem 2021;110:302-302. [doi: 10.1016/j.procbio.2021.10.011]
132. Sanchez-Campos N, Bernaldez-Sarabia J, Licea-Navarro AF. Conotoxin patenting trends in academia and industry. Mar Drugs 2022 Aug 19;20(8):531. [doi: 10.3390/md20080531] [Medline: 36005534]
133. Mendes CRA, Pappen E, Baeta AMC, et al. Patentometric analysis of the technological development of biotechnology for health in higher education institutions in Rio Grande do Sul. An Acad Bras Cienc 2024;96(1):e20231201. [doi: 10.1590/0001-3765202420231201] [Medline: 38451600]
134. Melo RLF, Neto FS, Dari DN, et al. A comprehensive review on enzyme-based biosensors: advanced analysis and emerging applications in nanomaterial-enzyme linkage. Int J Biol Macromol 2024 Apr;264(Pt 2):130817. [doi: 10.1016/j.ijbiomac.2024.130817] [Medline: 38479669]
135. Hachimi Alaoui C, Fatimi A. A 20-year patent review and innovation trends on hydrogel-based coatings used for medical device biofabrication. J Biomater Sci Polym Ed 2023 Jun;34(9):1255-1273. [doi: 10.1080/09205063.2022.2161777] [Medline: 36542377]
136. Tarasova EV, Luchnikova NA, Grishko VV, Ivshina IB. Actinomycetes as producers of biologically active terpenoids: current trends and patents. Pharmaceuticals (Basel) 2023 Jun 12;16(6):872. [doi: 10.3390/ph16060872] [Medline: 37375819]
137. Ragno L, Borboni A, Vannetti F, Amici C, Cusano N. Application of social robots in healthcare: review on characteristics, requirements, technical solutions. Sensors (Basel) 2023 Jul 31;23(15):6820. [doi: 10.3390/s23156820] [Medline: 37571603]
138. Chowdhury AR, Gargate G. The trends in CRISPR research: a patent and literature study with a focus on India. World Patent Information 2021 Jun;65:102038. [doi: 10.1016/j.wpi.2021.102038]
139. Pollock D, Peters MDJ, Khalil H, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 1;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]
140. Herman Muraro Gularte K, Alfredo Ruiz Vargas J, Paulo Javidi da Costa J, et al. Safeguarding the V2X pathways: exploring the cybersecurity landscape through systematic review. IEEE Access 2024;12:72871-72895. [doi: 10.1109/ACCESS.2024.3402946]
141. Difference between data mining and text mining. GeeksforGeeks. URL: https://www.geeksforgeeks.org/difference-between-data-mining-and-text-mining [accessed 2025-08-22]
142. Data transformation. Starburst. URL: https://www.starburst.io/data-glossary/data-transformation/ [accessed 2025-07-15]
143. Petryschuk P. What is network visualization? Auvik. 2024. URL: https://www.auvik.com/franklyit/blog/what-is-network-visualization/ [accessed 2025-08-22]
144. Hines P, Hiu Yu L, Guy RH, Brand A, Papaluca-Amati M. Scanning the horizon: a systematic literature review of methodologies. BMJ Open 2019 May 27;9(5):e026764. [doi: 10.1136/bmjopen-2018-026764] [Medline: 31133588]
145. Vignali V, Hines PA, Cruz AG, Ziętek B, Herold R. Health horizons: future trends and technologies from the European Medicines Agency's horizon scanning collaborations. Front Med (Lausanne) 2022;9:1064003. [doi: 10.3389/fmed.2022.1064003] [Medline: 36569125]
146. Global cancer burden growing, amidst mounting need for services. World Health Organization. 2024. URL: https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services [accessed 2025-07-15]
147. Naghavi M, Mestrovic T, Gray A, et al. Global burden associated with 85 pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet Infect Dis 2024 Aug;24(8):868-895. [doi: 10.1016/S1473-3099(24)00158-0]
148. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Proc Int AAAI Conf Web Soc Media 2009;3(1):361-362. [doi: 10.1609/icwsm.v3i1.13937]
149. van Eck NJP, Waltman L. VOS: a new method for visualizing similarities between objects. In: Decker R, Lenz HJ, editors. Advances in Data Analysis: Springer Berlin Heidelberg; 2007.
150. Derwent innovation: explanation on themescape. Clarivate. 2022. URL: https://support.clarivate.com/Patents/s/article/Derwent-Innovation-Explanation-on-ThemeScape?language=en_US [accessed 2025-07-15]
151. About pandas. Pandas. 2025. URL: https://pandas.pydata.org/about/ [accessed 2025-07-15]
152. Bubela T, Gold ER, Graff GD, Cahoy DR, Nicol D, Castle D. Patent landscaping for life sciences innovation: toward consistent and transparent practices. Nat Biotechnol 2013 Mar;31(3):202-206. [doi: 10.1038/nbt.2521] [Medline: 23471065]
153. Smith JA, Arshad Z, Trippe A, Collins GS, Brindley DA, Carr AJ. The reporting items for patent landscapes statement. Nat Biotechnol 2018 Nov 9;36(11):1043-1047. [doi: 10.1038/nbt.4291] [Medline: 30412195]

Abbreviations

IO: Innovation Observatory
NICE: National Institute for Health and Care Excellence

XSL•FO
RenderX

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SDMO: study type, data, methods, outcomes

Edited by TDA Cardoso; submitted 19.12.24; peer-reviewed by IC Lai, JK Timmis, M Pruski; revised version received 27.05.25; accepted 28.05.25; published 11.09.25.

Please cite as:
Garcia Gonzalez-Moral S, Pennock E, Ewedairo O, Green E, Elgey J, Mkwashi A
Horizon Scanning Methods for Health Care Technology Innovation Identification: Rapid Scoping Review of Patent Research Studies
Interact J Med Res 2025;14:e70323
URL: https://www.i-jmr.org/2025/1/e70323
doi:10.2196/70323

© Sonia Garcia Gonzalez-Moral, Erin Pennock, Olushola Ewedairo, Elizabeth Green, James Elgey, Andrew Mkwashi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 11.9.2025. 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

eHealth Literacy Interventions: Scoping Review

Yan Wang[1*], MSN; Yutian Niu[2*], MSN; Rongjing Xu[1], MSN; Qingqing Zhang[1], MSN; Shoumei Jia[1], PhD; Anni Wang[1], PhD

[1]School of Nursing, Fudan University, No.305 Fenglin Road, Xuhui District, Shanghai, China

[2]GuiYang Healthcare Vocational University, Guiyang, China

[*]these authors contributed equally

Corresponding Author:
Anni Wang, PhD
School of Nursing, Fudan University, No.305 Fenglin Road, Xuhui District, Shanghai, China

Abstract

Background: Electronic resources have become a predominant modality for health information dissemination in recent years. eHealth literacy (eHL) means individuals' competencies to effectively acquire and use health information from electronic sources. Enhancing eHL is thus essential to facilitate individuals' effective engagement with electronic resources and promote improved health management.

Objective: This scoping review aimed to synthesize the characteristics of eHL interventions, thereby providing a reference for future intervention strategies.

Methods: A comprehensive search of PubMed, Embase, Cochrane Library, Web of Science, ProQuest, CINAHL, CNKI, VIP, Wan Fang Data, and Sino Med limited to Chinese and English-language studies published before August 2024 was conducted. The interventional studies included had the explicit primary objective of enhancing eHL. We also incorporated studies that assessed eHL as a secondary outcome or mediator influencing health behaviors or clinical outcomes. All publications were required to provide publicly accessible complete datasets. We excluded conference abstracts and protocols. Academic theses and dissertations were included if they underwent institutional quality assurance through rigorous academic review processes and met predefined eligibility criteria.

Results: A total of 35 studies were included in this review. The most prevalent eHL interventions (12/35, 34%) were delivered via mobile apps and devices in various settings, including educational institutions, public spaces, health care facilities, and community centers. These interventions predominantly focused on enhancing information literacy, health literacy, and computer literacy across the 6 domains of eHL: traditional, health, information, scientific, media, and computer literacy. A majority of the interventions were conducted on a weekly basis (6/13, 46%) and had a duration of 24 weeks (6/35, 17%). However, 77% (27/35) of interventions did not assess long-term effects. The primary outcomes of eHL interventions encompassed perceived eHL, actual eHealth knowledge and skills, health literacy, health behavior, and clinical outcomes, with 86% (30/35) indicating positive effects. The eHealth Literacy Scale was the most frequently used assessment tool.

Conclusions: This study synthesizes the characteristics of eHL interventions. Current eHL interventions exhibit limitations in theoretical grounding, longitudinal tracking, and traditional or media literacy components. Overreliance on self-reported metrics constrains validity assessment. Future work should strengthen theoretical frameworks, integrate objective metrics, and enhance longitudinal designs.

(Interact J Med Res 2025;14:e69640) doi:10.2196/69640

KEYWORDS

eHealth literacy; intervention; scoping review; digital health; health literacy; eHL

Introduction

Background

With the advancement of eHealth technologies, individuals increasingly access information related to diseases, treatment options, medications, and healthy lifestyles through electronic resources. As of 2024, the number of internet health care users in China has reached 365 million, constituting approximately 26% of the total population [1]. This statistic highlights the public's reliance on eHealth resources and underscores the significant role that electronic resources play in the acquisition of health information. Nevertheless, the quality of eHealth information available is highly variable, necessitating that users possess specific skills to discern and evaluate its authenticity and reliability.

In this context, the concept of "eHealth literacy" (eHL) has emerged. Initially proposed by Norman and Skinner [2], eHL

XSL•FO
RenderX

refers to the capacity to access, comprehend, evaluate, and apply health-related information from electronic resources to address health-related issues. Building upon this concept, Norman and Skinner [2] further proposed the Lily Model. According to this model, eHL encompasses 6 key domains: traditional literacy, information literacy, media literacy, health literacy, scientific literacy, and computer literacy. These 6 domains are interrelated and together form the capacity for individuals to effectively use health information in the digital environment.

eHL serves as the foundation for the effective use of eHealth resources and plays a crucial role in encouraging healthy behaviors, maintaining and promoting health, and preventing diseases [3]. However, eHL levels among populations globally require urgent improvement. A survey conducted in Northwest Ethiopia [4] revealed that over half (235/423, 56%) of the participants exhibited eHL levels below the population average. Similarly, research conducted in Germany [5] indicated that approximately 42% of surveyed individuals demonstrated inadequate performance on eHL assessments, where cancer patients particularly faced challenges in effectively navigating eHealth information. Studies from China [6,7] further substantiate this trend, consistently reporting suboptimal eHL proficiency within the general population in China. In recent years, a diverse array of eHL interventions has emerged. Conducting a review and analysis of these interventions can provide a comprehensive understanding of their implementation methodologies, application contexts, and key characteristics. This endeavor facilitates the identification of prevalent trends and existing limitations, thereby establishing a robust scientific foundation for optimizing intervention designs. Current research predominantly focuses on comprehensive analyses of individual eHL levels, measurement tools, and the relationship between eHL and health behaviors [8-10]. A preliminary search on PubMed revealed only 3 reviews on eHL interventions: 2 reviews summarized the effects of eHL interventions on older adults, both noting positive impacts on their health and management but lacking detailed intervention references [11,12], while another review summarized the application of eHL models in digital health interventions, revealing that the Lily Model has not been fully applied across its various modules [13]. Despite numerous studies investigating the effectiveness and characteristics of eHL interventions, there remains a notable absence of systematic reviews that comprehensively summarize the diverse array of these interventions. A thorough review of eHL interventions is urgently warranted to guide future research efforts and inform practical applications.

Objectives

This study aimed to synthesize existing eHL interventions, concentrating on intervention formats, learning methods, delivery settings, components of the intervention, theory, duration, frequency, evaluation timelines of the intervention, primary outcomes and measures, and intervention effects, as well as commonly used eHL assessment tools. By identifying the characteristics and limitations of current research, this study aimed to provide a reference for health care professionals and researchers, facilitating a deeper understanding of the implementation patterns of eHL interventions and providing evidence-based guidance for future initiatives.

Methods

Overview

This scoping review was conducted and reported in strict accordance with the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) guidelines [14]. This study did not conduct a formal critical appraisal or methodological quality assessment, as the purpose of this scoping review was to identify and map the characteristics of eHL interventions rather than to draw conclusions regarding intervention effectiveness. This checklist can be found in Checklist 1.

Identifying the Research Question

This study seeks to address these questions: (1) What are the characteristics of eHL interventions? and (2) In which domains do eHL interventions exhibit limitations?

Eligibility Criteria

The inclusion criteria for this review were based on the "Population-Concept-Context" framework: (1) participants of any groups; (2) for concept, the primary aim of eHL interventions was to enhance eHL itself or to improve health behaviors or clinical outcomes through its enhancement. Furthermore, this review also included studies in which eHL was assessed as a secondary outcome, allowing for a comprehensive assessment of both the interventions' effectiveness and their impact on health outcomes; and (3) for context, potential settings for the implementation of eHL interventions encompass a range of environments, including educational institutions, residential settings, health care facilities, public libraries, and diverse digital platforms.

Furthermore, in alignment with the PRISMA-ScR guidelines [14] for scoping reviews, this study included degree theses and dissertations that were significantly pertinent to the research topic, offered detailed descriptions of interventions, and provided thorough methodological accounts, all of which had undergone institutional quality assurance through academic review processes. This inclusion aimed to mitigate publication bias and present a more comprehensive representation of the available evidence.

The exclusion criteria included (1) research proposals or conference titles; (2) studies written in non-English or non-Chinese; (3) to guarantee the integrability of the studies included and the robustness of the research outcomes, this study ruled out those with incomplete or missing raw data reports; and (4) unavailability of the full text.

Information Sources

A comprehensive search was conducted in the following databases: PubMed, Embase, Cochrane Library, Web of Science, ProQuest, CINAHL, CNKI, VIP, Wan Fang Data, and Sino Med.

Search Strategy

The search strategy incorporated Medical Subject Headings terms, free-text terms, and Boolean operators. Moreover, reference lists were reviewed to uncover any further relevant

XSL•FO
RenderX

studies. The final search was performed on August 9, 2024. The full search strategies for each database are available in Multimedia Appendix 1. English search terms included: "digital health literacy," "eHealth literacy," "e-Health literacy," "electronic health literacy," "telehealth literacy," "internet-based health literacy," "online health literacy," "health information literacy," "health information seeking," "health information searching," "intervention," "strateg*," "program*," "protocol," and "practice." Chinese search terms included: "eHealth Literacy," "digital health literacy," "intervention," "program," and "strategy." A specific example of how the search was conducted in the PubMed database includes:

#1 was "digital health literacy" [Title/Abstract] OR "eHealth literacy" [Title/Abstract] OR "e-Health literacy" [Title/Abstract] OR "electronic health literacy" [Title/Abstract] OR "telehealth literacy" [Title/Abstract] OR "internet-based health literacy" [Title/Abstract] OR "online health literacy" [Title/Abstract] OR "health information literacy" [Title/Abstract] OR "health information seeking" [Title/Abstract] OR "health information searching" [Title/Abstract];

#2 was "intervention" [Title/Abstract] OR "strateg*" [Title/Abstract] OR "program*" [Title/Abstract] OR "protocol" [Title/Abstract] OR "practice" [Title/Abstract] OR "trial" [Title/Abstract] OR "experiment*" [Title/Abstract] OR "therapy" [Title/Abstract];

#3 was #1 AND #2.

Selection of Study and Data Extraction

The retrieved records were imported into EndNote (version X9; Clarivate) for consolidation and duplicate removal. Two researchers (YW and YN), who were familiar with eHL interventions, independently screened the titles and abstracts based on inclusion and exclusion criteria. Next, the full texts of relevant studies were retrieved for further detailed screening. Any disagreements were resolved through discussion between the researchers, or if necessary, adjudicated by a third researcher (AW). For data extraction, the same two researchers (YW and YN) independently extracted information using a standardized form. The extracted data were then cross-checked, and disagreements were resolved through consensus discussion. In cases of uncertainty, a third researcher (AW) was consulted to make the final determination. Data were extracted on the following elements: (1) author, year of publication, country, study design, participants, sample size, and theory; (2) formats, learning methods, delivery setting and components of intervention, duration, frequency and evaluation timelines of intervention, primary outcomes and measures, intervention effects, and eHL assessment tools. A descriptive synthesis was conducted to categorize the characteristics of eHL interventions into structured tabulations (Multimedia Appendix 2 [15-49]).

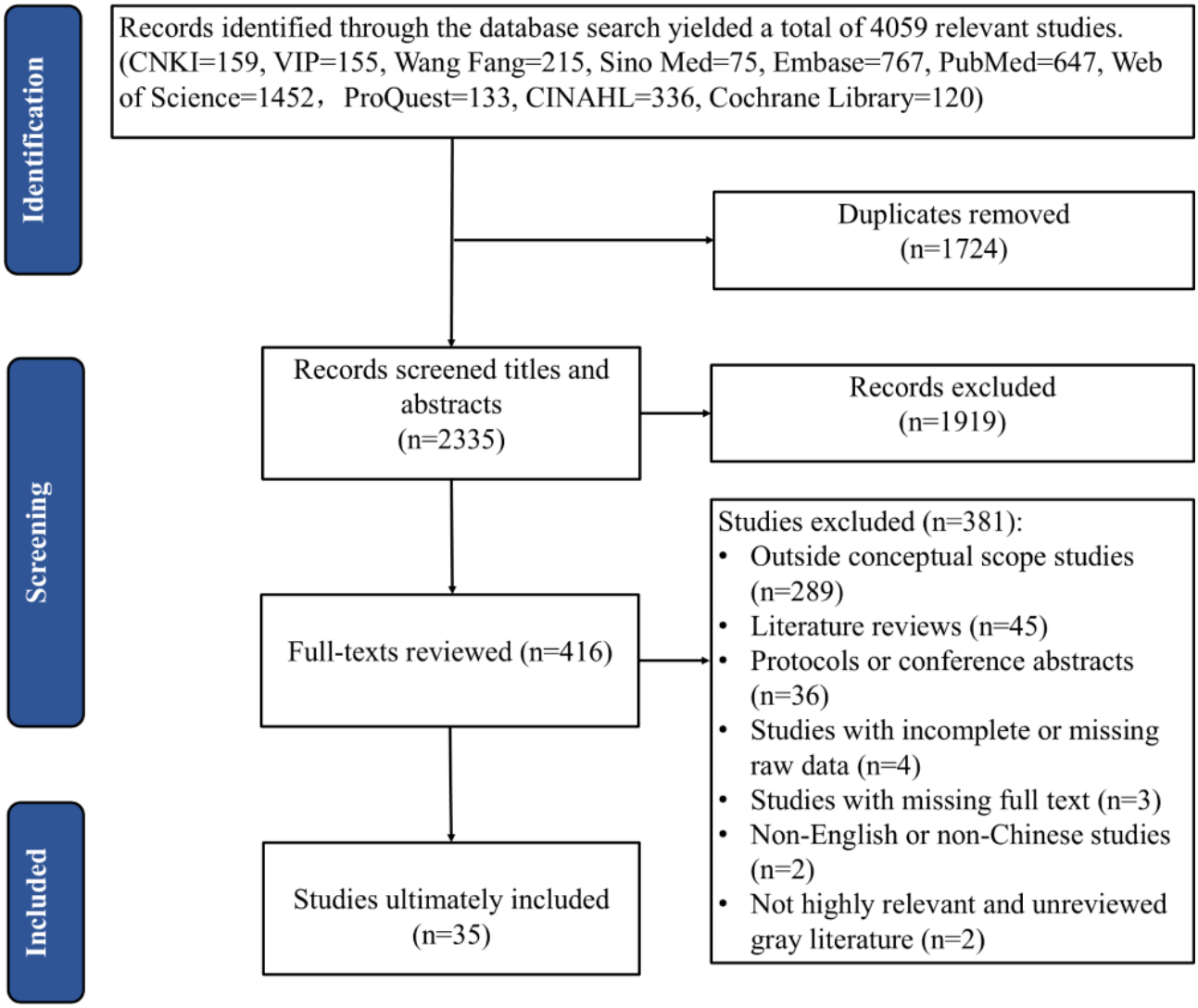
Results

Search Results

Initially, 4059 studies were retrieved. After the removal of 1724 duplicates, 1919 studies were excluded during title and abstract screening. Following a full-text review, an additional 381 studies were excluded, leading to the final inclusion of 35 studies. The detailed process is depicted in Figure 1.

XSL•FO
RenderX

Figure 1. PRISMA flow diagram of the screening process. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Basic Information of Included Studies

Of the 35 studies [15-49] selected, 9 studies [27-29,31,32,37-39,48] were in Chinese and 26 were in English. Of these, 9 studies [15,27-29,31,32,37,38,48] comprised master's or doctoral dissertations, with the three most represented countries being China, the United States, and South Korea. The study designs included 12 randomized controlled trials and 23 quasi-experimental studies, with sample sizes ranging from 26 to 1389 participants. The publication years spanned from 2005 to 2024, reflecting an exponential increase in research output over this period, culminating in a peak in 2023. Participants in eHL interventions predominantly included individuals with chronic diseases, older adults, students, and those affected by infectious diseases. Among these groups, patients with chronic diseases emerged as the most frequently targeted demographic, with 16 studies specifically addressing this population. Older adults constituted the second largest group, as indicated in 8 studies. Interventions directed at students were documented in 4 studies, while 3 studies focused on individuals living with HIV. Furthermore, eHL interventions have been implemented for other distinct populations, including pregnant women, adolescents with epilepsy and their parents, adults, families, and health care professionals. Comprehensive information regarding the included studies is presented in Table 1, while specific details of the interventions are provided in Multimedia Appendix 2 [15-49].

XSL•FO
RenderX

Table . Basic information of included studies.

Author	Year of publication	Country	Study design	Participants	Sample size	Theory
Norman [15]	2005	Canada	RCT[a]	Teenagers	1389	N/A[b]
Xie [16]	2011	United States	Quasi-experimental	Older people (aged ≥60 years)	124	N/A
Xie [17]	2011	United States	Quasi-experimental	Older people (aged ≥60 years)	146	N/A
Xie [18]	2012	United States	Quasi-experimental	Older people (aged ≥60 years)	146	N/A
Mitsuhashi [19]	2018	Japan	RCT	Adults (aged 20–59 years)	282	N/A
Carroll et al [20]	2019	United States	RCT	Persons living with HIV	360	Capability, Opportunity, Motivation, and Behavior model
Nokes and Reyes [21]	2019	United States	Quasi-experimental	Persons living with HIV	100	N/A
Nahm et al [22]	2019	United States	RCT	Older adults with chronic illnesses (aged ≥50 years)	272	Self-efficacy theory
Redfern et al [23]	2020	Australia	RCT	Patients with or at risk of cardiovascular disease	891	N/A
Lotto et al [24]	2020	Brazil	RCT	Parents of low socioeconomic preschoolers	208	N/A
Güvenc et al [25]	2020	Turkey	RCT	Youth with epilepsy and parents	70	N/A
Hyman et al [26]	2020	Canada	Quasi-experimental	Intermediate elementary students	245	N/A
Xu [27]	2021	China	Quasi-experimental	Patients with systemic lupus erythematosus	41	N/A
Jiang [28]	2021	China	Quasi-experimental	Patients with hypertension	60	Health as expanding consciousness theory
Guo [29]	2021	China	Quasi-experimental	High-risk population for stroke	229	Ecology of Health, Self-Determination Theory
Sanders et al [30]	2021	United States	RCT	Persons living with HIV	359	N/A
Zhou [31]	2022	China	Quasi-experimental	Postpartum women	152	N/A
Li [32]	2022	China	Quasi-experimental	Patients with nonalcoholic fatty liver disease	114	Behavior Change Wheel theory
Alt et al [33]	2022	Israel	Quasi-experimental	Undergraduate students (third year)	113	N/A
Chang et al [34]	2022	South Korea	Quasi-experimental	Older people (aged ≥65 years)	46	IMB[c] theory
Spindler et al [35]	2022	Denmark	RCT	Patients with heart failure	97	N/A
Parker et al [36]	2022	Australia	RCT	Patients with overweight and obesity attending primary care	215	N/A

XSL•FO
RenderX

Author	Year of publication	Country	Study design	Participants	Sample size	Theory
Liu [37]	2023	China	Quasi-experimental	Women at high risk of primary osteoporosis	80	Health Promotion Model
Yang [38]	2023	China	Quasi-experimental	Older hospitalized patients with coronary heart disease (aged ≥60 years)	82	IMB theory
Yu [39]	2023	China	RCT	Young and middle-aged patients with stroke (aged 18–54 years)	58	Transactional Model of eHealth Literacy
Dimitri et al [40]	2023	Australia	Quasi-experimental	Health care professionals	31	N/A
Son and Han [41]	2023	South Korea	Cluster randomized controlled trial	Multicultural families with young children	101	ADDIE[d] model
Roh and Won [42]	2023	South Korea	Quasi-experimental	Female college students	120	N/A
Son and Kim [43]	2023	South Korea	Quasi-experimental	Patients with heart failure	26	ADDIE model
Vazquez et al [44]	2023	United States	Quasi-experimental	Older people (aged ≥60 years)	466	N/A
Guo et al [45]	2023	China	Quasi-experimental	Patients with diabetes	132	N/A
Son et al [46]	2023	South Korea	Quasi-experimental	Patients with heart failure	100	IMB theory
Kang and You [47]	2023	South Korea	Quasi-experimental	Parents with children younger than 6 years of age	61	N/A
Hu [48]	2024	China	Quasi-experimental	Older patients with stroke (aged ≥60 years)	76	Information ecology theory
Intarakamhang et al [49]	2024	Thailand	Quasi-experimental	working-age people with risk factors for noncommunicable diseases	200	N/A

[a]RCT: randomized controlled trial.

[b]N/A: not applicable.

[c]IMB: Information-Motivation-Behavioral Skills.

[d]ADDIE: Analysis, Design, Development, Implementation, and Evaluation.

Characteristics of eHL Interventions

Formats of Intervention

The most prevalent of the included studies (12/35, 34%) used digital interventions through mobile apps or devices [19,20,23,24,28,35,36,40,41,43,46,49]. Website-based digital interventions were also relatively common, comprising 26% (9/35) of the included studies [15-18,22,25,42,44,47]. In contrast, a smaller proportion of studies (8/35, 23%) used face-to-face interventions [21,26,29,30,34,37,38,45]. Additionally, fewer studies (5/35, 14%) implemented remote digital interventions using video or communication programs [27,31-33,39], along with a combination of mobile apps and face-to-face interventions [48].

Learning Methods of Intervention

During the intervention process, various instructional methods were used, primarily categorized into interactive learning, independent learning, and a blended approach that combines both. Interactive learning emphasizes the active engagement of learners with the learning content, tasks, or problems, encompassing communication and collaboration among learners, as well as between learners and facilitators [50]. Within the 35 studies included, a substantial proportion (24/35, 69%) used independent learning methods [18,19,21-29,31,34-36,39-43,45-48]. The blended approach, integrating independent and interactive learning, was also common, comprising 23% (8/35) of the studies [16,17,20,32,37,38,44,49]. In contrast, research concentrated

XSL•FO
RenderX

on interactive learning was less prevalent, representing only 9% (3/35) of the studies [15,30,33].

Delivery Setting

Among the 35 studies included in this review, 14 studies [16-18,20,21,25,26,30,34,36,39,44,45,48] provided comprehensive reports on the intervention settings, which can be classified into 5 primary categories: educational institutions, public learning spaces, health care facilities, welfare centers, and recreational venues. Health care facilities were the most frequently reported settings, encompassing primary care clinics [20,30], general practice clinics [36], outpatient waiting areas [25,45], and hospital wards [39,48]. A total of 7 studies [20,25,30,36,39,45,48] were conducted within this category. Public learning spaces [16-18,44] and educational institutions [21,26] were the next most commonly reported settings. Furthermore, one study was conducted at a welfare center [34] and another at a recreational venue [44].

Components of Intervention

Overview

Using the 6 components of the Lily Model—information literacy, scientific literacy, health literacy, media literacy, traditional literacy, and computer literacy—the eHL intervention content of the included studies was systematically analyzed and summarized. The findings indicated that only one study [45] comprehensively addressed all 6 domains. The majority of studies (22/35, 63%) concentrated on a single domain, with health literacy being the most frequently targeted, as evidenced by 11 studies [20,23-25,28,35,36,40,46,47,49]. This was followed by information literacy and computer literacy, reported in 6 studies [15,16,21,27,32,43] and 5 studies [17,18,22,26,31], respectively.

Notably, some studies, although not encompassing all 6 domains, addressed multiple domains. For instance, one study [39] incorporated intervention content spanning 4 domains: information literacy, scientific literacy, health literacy, and computer literacy. Additionally, 3 studies [37,38,42] addressed three domains: information literacy, health literacy, and computer literacy. Furthermore, 8 studies focused on two domains, including combinations of information literacy and health literacy [19,29], health literacy and computer literacy [30,41], and information literacy and computer literacy [33,34,44,48].

Theory

In this study, less than half of the research (13/35, 37%) applied theories to guide the interventions [20,22,28,29,32,34,37-39,41,43,46,48]. Among these studies that applied theories, the majority of studies (7/35, 20%) drew on health behavior change fields [20,29,32,34,37,38,46], while relatively few studies incorporated theories in the fields of education and nursing training [28,29,41,43]. Additionally, apps in health informatics fields (2/35, 6%) [39,48] and health psychology fields (1/35, 3%) [22] were also limited.

The Information-Motivation-Behavioral Skills theory was the predominant framework in the domain of health behavior change, as evidenced by its application in 3 studies [34,38,46]. Additionally, the field also integrated various other theoretical frameworks, including the Behavior Change Wheel [32], the Health Promotion Model [37], the Capability, Opportunity, Motivation, and Behavior model [20], and the Ecology of Health [29]. Collectively, these theories illuminate the multidimensional aspects of behavior change, emphasizing the intricate interplay between individuals' intrinsic motivations and their external contexts.

The theories used in the domains of education and nursing training encompass the health as expanding consciousness theory [28], Self-Determination Theory [29], and the Analysis, Design, Development, Implementation, and Evaluation model [41,43]. These frameworks underscore the significance of individual cognition and self-identity within the educational process. In the realm of health informatics, the Transactional Model of eHealth Literacy [39] and information ecology theory [48] were prominently used. Furthermore, in health psychology, the self-efficacy theory [22] was applied.

Overall, the theories underpinning eHL interventions primarily focused on individual behavior change. These theories were designed to address various aspects of individual motivation, behavior, and the surrounding environment in developing intervention strategies.

Duration, Frequency, and Evaluation Timelines of Intervention

The duration of eHL interventions varied significantly, ranging from 2 weeks to 1 year. Most of the selected studies (6/35, 17%) [15,24,29,32,43,46] chose a duration of 24 weeks for the intervention. This was followed by 6 weeks [26,30,42,47,49] and 12 weeks [25,28,33,39,45], with 5 studies each. In addition, a small number of studies [17-19,22,27,31,34,40,44] reported intervention durations between 2 to 5 weeks, as well as 8 weeks [37,41], and 1 year [20,23,35,36].

The majority of studies (22/35, 63%) [15,16,20-23,25,26,28,29,32,33,35,36,38-41,43,45,46,48] did not provide specific details regarding the frequency of the interventions. However, among the 13 studies [17-19,24,27,30,31,34,37,42,44,47,49] that reported intervention frequency, most interventions [30,34,37,42,47,49] were conducted once weekly, followed by those occurring twice weekly [17,18,27,44]. Additionally, one study [24] reported an intervention frequency of once every 2 weeks, while another specified a daily intervention frequency [19]. Notably, in the study by Zhou [31], the frequency of community outreach was conducted daily, whereas the intervention delivered through instructional videos occurred once per week.

The evaluation timelines for outcome indicators primarily included before and after the intervention, as well as during the intervention and follow-up. Most of the studies (18/35, 51%) [16-19,23,25,27,28,31,33,37,38,40,42,43,47-49] chose to evaluate outcomes at the two critical time points of pre- and postintervention. Additionally, 8 studies [21,22,26,30,34,41,44,45] included follow-up assessments based on evaluations conducted before and after the intervention to assess its long-term effects. Among these studies, a majority (5/8, 63%) [21,22,30,41,45] of the interventions exhibited

XSL•FO
RenderX

sustained positive effects over the long term. However, 2 of these interventions [21,22] did not demonstrate statistically significant differences when compared to the control group during follow-up assessments. Furthermore, 3 additional studies [26,34,44] reported initial improvements immediately following the interventions; nonetheless, these effects did not persist throughout the follow-up periods, although the outcomes remained above baseline levels. A total of 9 studies [15,20,24,29,32,35,36,39,46] assessed outcome indicators both before and after the intervention, as well as during the intervention. Among these, only 1 study [35] reported significant positive effects in the first half of the intervention, but this effect was not evident during the remaining intervention period. Conversely, the other 8 studies [15,20,24,29,32,36,39,46] demonstrated improvements in primary outcomes compared to baseline, both during the entire intervention period and following its completion. However, in 2 of these studies [15,36], no statistically significant differences were observed between the intervention effects and the control group.

Primary Outcomes and Measures

The primary outcomes of eHL interventions encompassed perceived eHL, actual eHealth knowledge and skills (the specific eHealth competencies being tested), health literacy, and their impacts on health behaviors and clinical outcomes. These studies used multidimensional outcomes to comprehensively evaluate the effectiveness of eHL interventions.

The majority (18/35, 51%) of interventions had only one primary outcome. Among these, the most common was perceived eHL, reported in 10 studies [15,19,21,31,33-35,39,40,48], followed by health behaviors in 3 studies [23,41,46] (the adherence to guideline-recommended medications [≥80% of days covered for blood pressure and stain medications] [23], health-promoting lifestyle [41], and self-care behaviors [46]). Additionally, 2 studies [38,47] used health literacy as the primary outcome (health literacy on chronic diseases [38] and the knowledge of medication safety [47]), 2 studies [18,22] used actual eHealth knowledge and skills (computer and web knowledge [18], and patient portal knowledge [22]), and 1 study [20] exclusively used a clinical outcome (patient activation) as the primary outcome.

Additionally, 17 interventions [16,17,24-30,32,36,37,42-45,49] involved multiple primary outcomes. Among the 11 interventions with 2 primary outcomes, 9 of these [16,17,25-28,30,42,44] included perceived eHL as one of the primary outcomes, in combination with other aspects: actual eHealth knowledge and skills [16,17,27,44], health behaviors [26,28,42], or health literacy [25,30]. Of the remaining interventions, one [24] focused solely on health behaviors and clinical outcomes, while another [49] concentrated on health literacy alongside health behaviors. Among the 4 interventions with 3 primary outcomes, 2 studies [32,36] measured perceived eHL, health literacy, and clinical outcomes, 1 study [37] assessed perceived eHL, health behaviors, and clinical outcomes, while 1 study [43] assessed perceived eHL, health literacy, and health behaviors. For the 2 interventions with 4 primary outcomes, one [45] evaluated mobile eHL, actual eHealth knowledge and skills, health behaviors, and clinical outcomes, while the other [29] examined perceived eHL, health literacy, health behaviors, and clinical outcomes.

The majority of interventions (25/35, 71%) [15,19-23,25,26,28-31,33-35,38-43,46-49], measured primary outcomes via participant self-reports, while the remaining 10 studies used either objective tests alone [18] or a combination of self-reports and objective tests [16,17,24,27,32,36,37,44,45] for outcome measures.

Intervention Effects

Among the 22 studies [19,20,22,24,25,27-30,32,33,35,37-39,41,42,45-49] that adopted a between-group comparison approach instead of a within-participants pre-post comparison, eHL interventions significantly improved participants' perceived eHL [19,25,27-30,32,33,35,37,39,42,45,48] (eHL [19,25,27-30,32,33,37,39,42,48], literacy in using technology to process health information [35], and mobile eHL [45]), and health literacy [25,29,30,32,38,47] (HIV-related health literacy [30], health literacy on chronic diseases [32,38], epilepsy knowledge [25], knowledge of medication safety [47], and knowledge of stroke [29]). In terms of actual eHealth knowledge and skills, the interventions notably enhanced participants' knowledge and skills of mobile technology and the internet [45], and online health information discernment ability [27]. Regarding health behaviors, the interventions significantly promoted the participants' health behavior [29], self-management level [28], and health-promoting lifestyle [37,41]. Additionally, they encouraged children's dietary behavior of consuming sugar-free candy [24] and improved sufficient health behavior [49], as well as self-care behaviors [46]. For clinical outcomes, eHL interventions significantly reduced participants' liver fat content [32], perceived stress and depression [29], controlled the severity of early childhood caries [24], and improved bone health status [37], patient activation [20], and average blood glucose level over 3 months [45].

Among the 8 studies [16-18,26,34,40,43,44] using within-participants pre-post comparisons, eHL interventions significantly enhanced participants' perceived eHL [16,17,26,34,40,43,44]. These interventions also demonstrated significant improvements in participants' actual eHealth knowledge and skills, including website evaluation skills [16], as well as computer and web knowledge and skills [17,18,44]. Furthermore, the interventions significantly increased participants' knowledge of heart failure [43], while positively modifying and enhancing self-care behaviors [43].

Additionally, in 5 interventions [15,21,23,31,36], although the primary outcomes demonstrated some improvement postintervention compared to baseline, no statistically significant differences were observed between the intervention and control groups.

eHL Assessment Tools

Of the 33 studies [15-17,19-48] that measured eHL, the majority [15-17,19-34,36-39,41-48] used the eHealth Literacy Scale (eHEALS) for measurement. Developed by Norman and Skinner [2] in 2006, the eHEALS assesses individuals' self-perceived abilities to evaluate, locate, and apply eHealth information to

XSL•FO
RenderX

address health-related issues through eight concise questions. Owing to its broad applicability and the availability of multiple language versions, the scale has emerged as the most widely used tool for assessing eHL.

In addition to the eHEALS, one investigation used the eHealth Literacy Questionnaire (eHLQ) as a measure of eHL [35]. Developed by Kayser et al [51] in 2018, the eHLQ consists of a self-administered questionnaire comprising 35 items designed to evaluate several dimensions of eHL. These dimensions include the capacity to process health information via technology, understanding of health concepts and terminology, active participation in digital services, perceived feelings of safety and control, proactive engagement with digital services, and access to effective digital resources that are tailored to meet individual needs. The validity of this instrument has been confirmed across numerous countries.

Furthermore, Dimitri et al [40] did not use any standardized instruments; rather, a custom-designed questionnaire was developed specifically to align with the objectives of the study in evaluating eHL, including the definition of eHL and the digital tools for improving the detection of growth disorders.

In summary, approximately 97% (32/33) of the studies used validated or standardized instruments for measuring eHL, thereby enhancing the reliability and validity of the research findings, as well as improving the comparability and reproducibility of the results.

Discussion

Principal Findings

This study synthesized and analyzed the characteristics of eHL interventions, including intervention formats, learning methods of intervention, delivery settings, components of the intervention, theoretical foundations, intervention duration, frequency, evaluation timelines of the intervention, primary outcomes and measures, intervention effects, and eHL measurement tools. The findings provided a detailed reference for the design and implementation of future intervention studies in this domain.

The formats of eHL interventions exhibited considerable diversity across various delivery settings. For instance, 2 studies [21,26] within educational environments used face-to-face intervention methods, likely due to the availability of resources and an environment conducive to in-person interactions. Conversely, 4 studies [16-18,44] conducted in public libraries used web-based interventions, leveraging the open and quiet nature of these settings, which are particularly well-suited for such approaches. In health care institutions, the predominant modalities for intervention delivery included mobile apps or devices [20,36], websites [25], and video communication programs [39]. This preference for digital interventions in health care settings may be attributed to constraints related to equipment and physical space, rendering digital solutions a more practical and scalable alternative. Among the studies included, the predominant format of eHL interventions was digital delivery via mobile apps or devices, which aligns with the findings of Bashir et al [52]. This preference can be attributed to the convenience of this approach, which facilitates intervention monitoring and enables broader scalability beyond specific study samples. Given that eHL emphasizes the capacity to locate, access, understand, evaluate, and use health information through electronic resources to address health-related issues, interventions using mobile apps or devices gain wider adoption. However, the majority of these studies [19,20,23,24,28,35,36,40,41,43,46,49] originated in high-resource countries, with merely 3 studies [24,28,49] conducted in resource-limited countries. Notably, eHL interventions in resource-limited countries predominantly adopt face-to-face offline or remote delivery via video or communication programs. This disparity underscores the persistent global digital divide: pervasive high-speed internet access and elevated smartphone penetration in high-resource countries establish an enabling environment for mobile app or device-dependent eHL interventions. Conversely, most resource-limited countries, with China constituting a notable exception, continue to face persistent challenges such as fragmented network coverage and prohibitive device costs [53]. In such settings, in-person sessions or asynchronous delivery methods could become pragmatic alternatives to mitigate access barriers. Therefore, eHL intervention designs should be context-specific, prioritizing infrastructure-appropriate implementation frameworks. Finally, although independent learning was the most commonly used intervention method in this study, a meta-analysis [54] demonstrated no significant association between the learning method and the intervention effect. This finding implies that the learning method may not be the primary determinant influencing the efficacy of the intervention.

This study, guided by the 6 domains of the Lily Model, conducted a thorough analysis of the content within eHL interventions and identified variations in the focal areas across different initiatives. Overall, these interventions predominantly emphasized health literacy, information literacy, and computer literacy. This observation aligns with Benny et al's findings [13]. Several factors may explain this trend: first, within the fields of health care and nursing, health literacy is widely recognized as a crucial determinant of patients' ability to manage their health and prevent diseases [55]. Many eHL interventions aim to enhance individuals' eHL with the overarching goal of improving their overall health, which naturally prioritizes health literacy. Second, with the increasing diversification of health information sources in recent years, the significance of information literacy and computer literacy has also escalated. Notably, only one study [45] incorporated intervention components that addressed all 6 domains of the Lily Model, while traditional literacy and media literacy received the least coverage. This finding aligns with a recent review by Benny et al [13], indicating that the limited focus on these domains constrains the generalizability of the efficacy of interventions. Furthermore, this suggests that eHL research and practice demonstrate relative deficiencies in attention and investment in traditional literacy and media literacy domains. We recommend emphasizing content integrity when implementing the Lily Model to ensure intervention comprehensiveness and efficacy generalizability [13].

XSL•FO
RenderX

The theory plays a crucial role in intervention strategies, which can provide guidance for intervention design, implementation, and evaluation. Brownell [56] emphasized that an appropriate theory would help identify risk factors for specific diseases and guide the development of interventions. A systematic review by Wong and Morrison [57] revealed that a lack of theoretical support frequently resulted in inadequate systematic and targeted approaches, potentially leading to content that fails to comprehensively address the needs of the target population. However, this study demonstrated that a significant majority (22/35, 63%) of included studies failed to incorporate theoretical frameworks in their interventions, which may limit the depth of understanding and analysis concerning the mechanisms by which certain interventions may be effective or ineffective. This finding aligns with the results of Chang et al [58], which highlighted the lack of theoretical support in the formulation and implementation of eHL interventions. Notably, this finding contrasts with the results of Dong et al [12], who reported that more than half of the 7 studies included in their review used theory. This discrepancy may be attributed to the smaller number of studies analyzed in their review. Furthermore, among the 13 interventions that incorporated theories, the majority emerged from the domain of health behavior change. These theories aim to design effective interventions by elucidating individuals' behavioral motivations, capabilities, and social environments, thereby enhancing individuals' eHL. Therefore, future eHL interventions should place greater emphasis on the application of theoretical frameworks, enhancing the scientific rigor and effectiveness of interventions.

In terms of the duration, frequency, and evaluation timelines of eHL interventions, this study found that the most commonly used intervention duration was 24 weeks, followed by shorter durations of 6 and 12 weeks. This finding is similar to the results of Dong et al [12], who reported that interventions lasting 4 weeks or more significantly improved individuals' eHL. The limited effectiveness of shorter interventions may be attributed to insufficient time for participants to fully absorb and apply the knowledge acquired, thereby affecting the overall impact of the intervention. In contrast, relatively longer intervention durations allow participants to repeatedly apply the learned skills in practice, thereby enhancing their confidence and capabilities, which in turn improves their eHL. In the design of intervention studies, the frequency of the intervention may significantly influence its effectiveness [59]. However, we found that most studies (22/35, 63%) omitted detailed information regarding intervention frequency, limiting understanding of the relationship between intervention effects and frequency. Among those studies that did specify frequency, the majority implemented interventions on a weekly basis. This finding provides a valuable reference for future eHL interventions. Future research should explore the specific effects of varying intervention frequencies on outcomes and assess whether the effectiveness of intervention frequencies differs across various health issues and target populations. Furthermore, the majority (27/35, 77%) of interventions did not conduct follow-up assessments of the intervention effects. This limitation may hinder researchers' ability to fully comprehend the true effects of the interventions, consequently impacting evaluations of their long-term feasibility and effectiveness [60]. Among the studies that addressed the sustainability of intervention effects, a majority (5/8, 63%) of the interventions reported that the interventions yielded long-term positive outcomes. Notably, Nokes et al [21] and Nahm et al [22] reported no statistically significant between-group differences during follow-up, suggesting that improvements in primary outcomes may not be solely attributable to the intervention effects but could also be influenced by various other factors. Therefore, future research should continue to investigate the factors that contribute to the long-term effectiveness of eHL interventions, including various aspects such as the intervention's form, duration, and frequency, as well as the characteristics of the target population.

The primary outcomes of eHL interventions encompass multiple dimensions, including perceived eHL, health literacy, actual eHealth knowledge and skills, health behaviors, and clinical outcomes. Notably, single primary outcomes were the most prevalent, comprising 51% (18/35) of the total, with perceived eHL being the most frequently reported at 56% (10/18). Among studies measuring multiple primary outcomes, the majority (11/17, 65%) focused on 2 types, with the combination of perceived eHL and actual eHealth knowledge and skills being the most prevalent (4/11, 36%). This finding is consistent with the results of a recent systematic review [54]. However, Watkins and Xie [11] found that most studies relied primarily on health outcomes as the main measure to evaluate eHL interventions. Compared to health outcome measures, composite indicators of perceived eHL and actual eHealth knowledge and skills can more directly and comprehensively focus on the core elements of eHL itself. This approach avoids the complexities and uncertainties that may arise from indirectly measuring eHL through health outcomes [61]. Regarding assessment tools for eHL, 3 instruments were identified: the eHEALS, the eHLQ, and custom-designed questionnaires. The findings indicated that 97% (32/33) of researchers used validated standardized self-report tools, with eHEALS being the predominant instrument. This trend aligns with the results reported by Kim et al [9]. One possible explanation for this preference is that eHEALS is more concise than the eHLQ, which contains 35 items and is available in multiple language versions. The excessive number of items in the eHLQ may lead to participant fatigue, potentially compromising the validity of the information collected. In contrast, while the custom-designed questionnaire developed by Dimitri et al [40] consists of only 10 questions, its validity has not been established, and it lacks generalizability. The results of this study showed that the majority of eHL interventions (30/35, 86%) achieved positive effects. However, notably, the assessment of primary outcome measures in most studies (25/35, 71%) relied solely on self-reported data from participants, which may compromise the reliability of findings and potentially lead to an overestimation of intervention effects [62]. Moreover, over half (23/35, 66%) were not randomized controlled trials, resulting in poor control of confounding factors. In light of this, future interventions should optimize research design and methods of outcome assessment.

Limitations

The limitations of this review are primarily manifested in several key areas. Due to potential constraints in cross-linguistic data extraction and analytical precision, this study exclusively

XSL•FO
RenderX

included English and Chinese literature. While this approach reinforces methodological rigor, future research could enhance evidence synthesis through multilingual collaboration. Nevertheless, given that most publications in this field are in English, our findings remain representative of the prevailing evidence. Furthermore, during literature screening and data extraction, this study did not formally calculate interrater reliability. While individual interpretation differences are inevitable, we implemented rigorous measures to minimize subjective bias, including predefined inclusion criteria to standardize decision-making, dual independent screening and extraction by two researchers, consensus resolution for initial discrepancies, and third-party arbitration for unresolved cases, with documented disagreement points. Our records indicate minimal divergence, as only two studies (both grey literature) necessitated arbitration during screening. Following the third-reviewer discussion, these studies were excluded from analysis.

Conclusions

This study reviewed eHL interventions, highlighting key characteristics and limitations. Predominant frameworks used health behavior theory, with mobile apps or devices serving as the primary intervention format; a 24-week duration was most frequent. Setting selection reflected contextual adaptability: educational and health care institutions favored in-person interactions, whereas public libraries and health care systems preferred digital interventions. Challenges include inadequate theoretical guidance, a lack of long-term follow-up, and overreliance on self-reported measures. Content emphasizes health literacy, information literacy, and computer literacy, but insufficient traditional literacy and media literacy coverage may limit effect generalizability. Although most interventions reported positive outcomes, nonrandomized designs and subjective assessments weaken the strength of the evidence. We recommend that future research prioritize enhancing theoretical integration, adopting more objective assessment methodologies, and improving the long-term follow-up of intervention effects to bolster the scientific rigor and practical value of eHL interventions.

Acknowledgments

This study was funded by the National Social Science Foundation (24CSH086). The funders had no role in the study design, data collection, analysis, interpretation, manuscript writing, or decision to publish. No generative artificial intelligence tools were used in the writing of this manuscript.

Authors' Contributions

YW, YN, RX, QZ, SJ, and AW conceptualized and designed the study; YN, RX, and QZ conducted the literature search and reference management. YW and YN selected relevant studies and performed data collection and assembly. YW and YN conducted data analysis and result interpretation. YW and AW drafted the manuscript. AW and SJ jointly provided supervision and timeline control of the study as senior authors. All authors will participate in the revision of the manuscript and approve the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full search strategies.
[DOC File, 44 KB - i-jmr_v14i1e69640_app1.doc]

Multimedia Appendix 2
The specific details of the interventions.
[XLSX File, 24 KB - i-jmr_v14i1e69640_app2.xlsx]

Checklist 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist
[DOC File, 55 KB - i-jmr_v14i1e69640_app3.doc]

References

1. Statistical report on the development of internet in china, 54th edition. Center CINI. 2024. URL: https://www.199it.com/archives/1718242.html [accessed 2024-11-05]
2. 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]
3. Brørs G, Norman CD, Norekvål TM. Accelerated importance of eHealth literacy in the COVID-19 outbreak and beyond. Eur J Cardiovasc Nurs 2020 Aug;19(6):458-461. [doi: 10.1177/1474515120941307] [Medline: 32667217]

XSL•FO
RenderX

4. Shiferaw KB, Tilahun BC, Endehabtu BF, Gullslett MK, Mengiste SA. E-health literacy and associated factors among chronic patients in a low-income country: a cross-sectional survey. BMC Med Inform Decis Mak 2020 Aug 6;20(1):181. [doi: 10.1186/s12911-020-01202-1] [Medline: 32762745]
5. Heiman H, Keinki C, Huebner J, Working Group Prevention and Integrative Oncology of the German Cancer Society. EHealth literacy in patients with cancer and their usage of web-based information. J Cancer Res Clin Oncol 2018 Sep;144(9):1843-1850. [doi: 10.1007/s00432-018-2703-8] [Medline: 30006768]
6. Liu Q, Huang R, Xia L, et al. eHealth literacy of chinese residents during the coronavirus disease 2019 pandemic: a cross-sectional survey. Comput Inform Nurs 2023 May 1;41(5):292-299. [doi: 10.1097/CIN.0000000000000921] [Medline: 35470296]
7. Dandan L, Yexiang Y, Yue Y, Yuguang Z, Li B. Research progress on eHealth literacy and influencing factors of residents in our country. Mhealth 2023;20:159-161. [doi: 10.16659/j.cnki.1672-5654.2023.06.159]
8. Verma R, Saldanha C, Ellis U, Sattar S, Haase KR. eHealth literacy among older adults living with cancer and their caregivers: a scoping review. J Geriatr Oncol 2022 Jun;13(5):555-562. [doi: 10.1016/j.jgo.2021.11.008]
9. Kim K, Shin S, Kim S, Lee E. The relation between eHealth literacy and health-related behaviors: systematic review and meta-analysis. J Med Internet Res 2023 Jan 30;25:e40778. [doi: 10.2196/40778] [Medline: 36716080]
10. 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]
11. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225. [doi: 10.2196/jmir.3318] [Medline: 25386719]
12. Dong Q, Liu T, Liu R, Yang H, Liu C. Effectiveness of digital health literacy interventions in older adults: single-arm meta-analysis. J Med Internet Res 2023 Jun 28;25:e48166. [doi: 10.2196/48166] [Medline: 37379077]
13. El Benny M, Kabakian-Khasholian T, El-Jardali F, Bardus M. Application of the eHealth literacy model in digital health interventions: scoping review. J Med Internet Res 2021 Jun 3;23(6):e23473. [doi: 10.2196/23473] [Medline: 34081023]
14. 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]
15. Norman CDZ. The web of influence: evaluating the impact of internet interventions on adolescent smoking cessation and ehealth literacy [Dissertation]. : University of Toronto; 2005.
16. 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]
17. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res ;13(4):e90. [doi: 10.2196/jmir.1880]
18. Xie B. Improving older adults' e-health literacy through computer training using NIH online resources. Libr Inf Sci Res 2012 Jan;34(1):63-71. [doi: 10.1016/j.lisr.2011.07.006]
19. Mitsuhashi T. Effects of two-week e-learning on eHealth literacy: a randomized controlled trial of Japanese internet users. PeerJ 2018;6:e5251. [doi: 10.7717/peerj.5251] [Medline: 30013857]
20. Carroll JK, Tobin JN, Luque A, et al. "Get Ready and Empowered About Treatment" (GREAT) study: a pragmatic randomized controlled trial of activation in persons living with HIV. J Gen Intern Med 2019 Sep;34(9):1782-1789. [doi: 10.1007/s11606-019-05102-7] [Medline: 31240605]
21. Nokes KM, Reyes DM. Do brief educational sessions increase electronic health literacy of low-income persons living with HIV/AIDS? Comput Inform Nurs 2019 Jun;37(6):315-320. [doi: 10.1097/CIN.0000000000000515] [Medline: 30870187]
22. Nahm ES, Zhu S, Bellantoni M, et al. The effects of a theory-based patient portal e-learning program for older adults with chronic illnesses. Telemed J E Health 2019 Oct;25(10):940-951. [doi: 10.1089/tmj.2018.0184] [Medline: 30431393]
23. Redfern J, Coorey G, Mulley J, et al. A digital health intervention for cardiovascular disease management in primary care (CONNECT) randomized controlled trial. NPJ Digit Med 2020;3(1):117. [doi: 10.1038/s41746-020-00325-z] [Medline: 32964140]
24. Lotto M, Strieder AP, Ayala Aguirre PE, et al. Parental-oriented educational mobile messages to aid in the control of early childhood caries in low socioeconomic children: a randomized controlled trial. J Dent (Shiraz) 2020 Oct;101:103456. [doi: 10.1016/j.jdent.2020.103456]
25. Tutar Güven Ş, İşler Dalgiç A, Duman Ö. Evaluation of the efficiency of the web-based epilepsy education program (WEEP) for youth with epilepsy and parents: a randomized controlled trial. Epilepsy Behav 2020 Oct;111:107142. [doi: 10.1016/j.yebeh.2020.107142] [Medline: 31766003]
26. Hyman A, Stewart K, Jamin AM, et al. Testing a school-based program to promote digital health literacy and healthy lifestyle behaviours in intermediate elementary students: the Learning for Life program. Prev Med Rep 2020 Sep;19:101149. [doi: 10.1016/j.pmedr.2020.101149] [Medline: 32670779]
27. Xu L. Construction of Intervention Program for Improving the Ability of Online Health Information Identification in Patients with Systemic Lupus Erythematosus [Master's Thesis]. : Anhui Medical University; 2021.
28. Jiang Y. Establishment and Application of Community Hypertension Patients Active Health Management Platform [Master's Thesis]. : University of South China; 2021.

XSL•FO
RenderX

29. Guo L. Construction and Application of Health Management Model for People at High Risk of Stroke Based on Health Ecology Theory [Doctoral Dissertation]. : The First Afiliated Hospital of Zhengzhou University; 2021 URL: https://link.cnki.net/doi/10.27466/d.cnki.gzzdu.2021.005375
30. Sanders M, Tobin JN, Cassells A, et al. Can a brief peer-led group training intervention improve health literacy in persons living with HIV? Results from a randomized controlled trial. Patient Educ Couns 2021 May;104(5):1176-1182. [doi: 10.1016/j.pec.2020.10.031] [Medline: 33221117]
31. Zhou M. Research on Influencing Factors of Maternal Mobile Health Service Use Behavior and Community Intervention Practice [Master's Thesis]. : Shenzhen University; 2022.
32. Li Y. Application of Intervention Program Based on Behavioral Change Wheel Theory in Patients with Non-Alcoholic Fattyliver Disease [Master's Thesis]. : Zhengzhou University; 2022.
33. Alt D, Naamati-Schneider L, Meirovich A. Future problem-solving practiced during COVID-19: implications for health management students' E-Health literacy identity. Front Psychol 2022;13:829243. [doi: 10.3389/fpsyg.2022.829243] [Medline: 35250771]
34. Chang SJ, Lee KE, Yang E, Ryu H. Evaluating a theory-based intervention for improving eHealth literacy in older adults: a single group, pretest-posttest design. BMC Geriatr 2022 Nov 29;22(1):918. [doi: 10.1186/s12877-022-03545-y] [Medline: 36451127]
35. Spindler H, Dyrvig AK, Schacksen CS, et al. Increased motivation for and use of digital services in heart failure patients participating in a telerehabilitation program: a randomized controlled trial. Mhealth 2022;8:25. [doi: 10.21037/mhealth-21-56] [Medline: 35928510]
36. Parker SM, Barr M, Stocks N, et al. Preventing chronic disease in overweight and obese patients with low health literacy using eHealth and teamwork in primary healthcare (HeLP-GP): a cluster randomised controlled trial. BMJ Open 2022 Nov 30;12(11):e060393. [doi: 10.1136/bmjopen-2021-060393] [Medline: 36450426]
37. Liu H. The Effect of Nursing Intervention Program Based on Health Promotion Model in Women at High Risk of Primary Osteoporosis [Master's Thesis]. : University of South China; 2023.
38. Yang Z. Health Literacy Interventions Based on IMB Models in Old Age: Application Research in Inpatient Patients with Coronary Heart Disease [Master's Thesis]. : Jiangsu University; 2023.
39. Yu J, Kang M, Luo Y, et al. Construction and application of an interactive model intervention for electronic health literacy in young and middle-aged stroke patients. J Nurs 2023;38:1-4+14. [doi: 10.3870/j.issn.1001-4152.2023.24.001]
40. Dimitri P, Fernandez-Luque L, Koledova E, Malwade S, Syed-Abdul S. Accelerating digital health literacy for the treatment of growth disorders: the impact of a massive open online course. Front Public Health 2023;11:1043584. [doi: 10.3389/fpubh.2023.1043584] [Medline: 37143968]
41. Son H, Han G. Development and evaluation of a tailored mHealth parenting program for multicultural families: a three-arm cluster randomized controlled trial. Front Public Health 2023;11:1182310. [doi: 10.3389/fpubh.2023.1182310] [Medline: 37886055]
42. Roh M, Won Y. Impact of online-delivered eHealth literacy intervention on eHealth literacy and health behavior outcomes among female college students during COVID-19. Int J Environ Res Public Health 2023 Jan 22;20(3):2044. [doi: 10.3390/ijerph20032044] [Medline: 36767409]
43. Son YJ, Kim EY. The effectiveness and usability of a novel mobile phone-based self-care intervention for patients with heart failure: a mixed-methods pilot study. Eur J Cardiovasc Nurs 2023 Apr 12;22(3):254-263. [doi: 10.1093/eurjcn/zvac077]
44. Vazquez CE, Xie B, Shiroma K, Charness N. Individualistic versus collaborative learning in an eHealth literacy intervention for older adults: quasi-experimental study. JMIR Aging 2023 Feb 9;6:e41809. [doi: 10.2196/41809] [Medline: 36757773]
45. Guo SHM, Lin JL, Hsing HC, Lee CC, Chuang SM. The effect of mobile eHealth education to improve knowledge, skills, self-care, and mobile ehealth literacies among patients with diabetes: development and evaluation study. J Med Internet Res 2023 Dec 6;25:e42497. [doi: 10.2196/42497] [Medline: 38055321]
46. Son YJ, Kim SW, Lee WS, et al. The effects of a 24-week interactive text message-based mobile health intervention for enhancing self-care behaviours of patients with heart failure: a quasi-experimental study. Nurs Open 2023 Sep;10(9):6309-6319. [doi: 10.1002/nop2.1878] [Medline: 37313589]
47. Kang NG, You MA. The effects of a smartphone-based education program designed to help mothers safely administer medication to their children. Comput Inform Nurs 2023 Feb 1;41(2):77-85. [doi: 10.1097/CIN.0000000000000859] [Medline: 35314588]
48. Hu Y. Construction and Empirical Study of Electronic Healthliteracy Intervention Program for Elderly Patients Withchronic Diseases— a Case Study of Stroke Patients [Master's Thesis]. : Nanjing University of Chinese Medicine; 2024.
49. Intarakamhang U, Sriprasertpap K, Chiangkhong A, Srisawasdi N, Wongchan S, Boocha P. Effects of digital health literacy program on sufficient health behavior among thai working-age people with risk factors for noncommunicable diseases. HLRP: Health Literacy Research and Practice 2024 Apr;8(2). [doi: 10.3928/24748307-20240520-01]
50. Berge ZL. Interaction in post-secondary web-based learning. Educ Technol Arch 1999;39:5-11. [doi: 10.13016/M2DR5I-7BUE]

XSL•FO
RenderX

51. 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]
52. Bashir AZ, Yetman A, Wehrmann M. Technological interventions to implement prevention and health promotion in cardiovascular patients. Healthcare (Basel) 2024 Oct 16;12(20):2055. [doi: 10.3390/healthcare12202055] [Medline: 39451470]
53. Global Internet use continues to rise but disparities remain, especially in low-income regions. ITU. 2024. URL: https://www.itu.int/en/mediacentre/Pages/PR-2024-11-27-facts-and-figures.aspx [accessed 2025-06-22]
54. Barbati C, Maranesi E, Giammarchi C, et al. Effectiveness of eHealth literacy interventions: a systematic review and meta-analysis of experimental studies. BMC Public Health 2025 Jan 23;25(1):288. [doi: 10.1186/s12889-025-21354-x] [Medline: 39849354]
55. Walters R, Leslie SJ, Polson R, Cusack T, Gorely T. Establishing the efficacy of interventions to improve health literacy and health behaviours: a systematic review. BMC Public Health 2020 Dec;20(1). [doi: 10.1186/s12889-020-08991-0]
56. Brownell P. Trauma theory and abuse, neglect and violence across the life course. J Interpers Violence 2024 Oct;39(19-20):4041-4064. [doi: 10.1177/08862605241264542] [Medline: 39254267]
57. Wong YT, Morrison SC. Health promotion models for improving footcare in older adults: a scoping review. Gerontology 2024;70(8):785-800. [doi: 10.1159/000538868] [Medline: 38636462]
58. Chang SJ, Jang SJ, Lee H, Kim H. Building on evidence to improve eHealth literacy in older adults: a systematic review. Comput Inform Nurs 2021 May 1;39(5):241-247. [doi: 10.1097/CIN.0000000000000674] [Medline: 32925185]
59. Lu SC, Xu M, Wang M, et al. Effectiveness and minimum effective dose of app-based mobile health interventions for anxiety and depression symptom reduction: systematic review and meta-analysis. JMIR Ment Health 2022 Sep 7;9(9):e39454. [doi: 10.2196/39454] [Medline: 36069841]
60. Butcher NJ, Monsour A, Mew EJ, et al. Guidelines for reporting outcomes in trial reports: the CONSORT-outcomes 2022 extension. JAMA 2022 Dec 13;328(22):2252-2264. [doi: 10.1001/jama.2022.21022] [Medline: 36511921]
61. 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]
62. Prince SA, Adamo KB, Hamel M, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5(1):56. [doi: 10.1186/1479-5868-5-56]

Abbreviations

eHEALS: eHealth Literacy Scale
eHL: eHealth literacy
eHLQ: eHealth Literacy Questionnaire
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

Edited by TDA Cardoso; submitted 04.12.24; peer-reviewed by A Azizi, E Neter, S Thawillarp; revised version received 27.06.25; accepted 27.06.25; published 22.08.25.

Please cite as:
Wang Y, Niu Y, Xu R, Zhang Q, Jia S, Wang A
eHealth Literacy Interventions: Scoping Review
Interact J Med Res 2025;14:e69640
URL: https://www.i-jmr.org/2025/1/e69640
doi:10.2196/69640

©Yan Wang, Yutian Niu, Rongjing Xu, Qingqing Zhang, Shoumei Jia, Anni Wang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.8.2025. 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

Safety and Efficacy of Aspirin and Indobufen in the Treatment of Atherosclerotic Diseases: Systematic Review and Meta-Analysis

Wenhao Pan*, BSc; Linger Guan*, BSc; Haicheng Zhang, MD

Department of Cardiology, Peking University People's Hospital, 11 Xizhimen S St, Xicheng District, Beijing, China

*these authors contributed equally

Corresponding Author:
Haicheng Zhang, MD
Department of Cardiology, Peking University People's Hospital, 11 Xizhimen S St, Xicheng District, Beijing, China

Abstract

Background: The pathogenesis of atherosclerotic thrombosis primarily involves platelet activation and aggregation, making antiplatelet therapy the cornerstone of treatment for such diseases.

Objective: This meta-analysis aimed to compare the safety and efficacy of aspirin and indobufen in antiplatelet therapy for patients with atherosclerotic diseases.

Methods: We searched the Cochrane Library, PubMed, Embase, Web of Science, and the Chinese Wanfang databases. The literature was screened according to predefined inclusion and exclusion criteria. Risk ratio (RR) was used to assess the magnitude of risk associated with exposure and our inclusion criteria are as follows: (1) the study population comprised adults (aged 18 years and older) with coronary heart disease caused by coronary artery atherosclerosis or stroke caused by intracranial atherosclerosis; (2) the intervention was represented by indobufen in the study groups versus aspirin in the control groups; (3) the primary outcome was the incidence of major adverse cardiovascular and cerebrovascular events or any bleeding or Bleeding Academic Research Consortium 2/3/5 bleeding; (4) the secondary outcomes were cardiovascular death, myocardial infarction, and ischemic stroke; adverse cardiovascular events such as coronary thrombus reformation, heart failure, myocardial infarction, stroke, angina pectoris, and cardiovascular death; stroke; myocardial infarction; and cardiovascular death; and (5) the studies were randomized clinical trials with either crossover or parallel designs or prospective observational trials.

Results: Eighteen trials with a total of 12,981 patients were included in this study. Compared to aspirin, indobufen reduced the risk of (1) bleeding events (RR 0.54, 95% CI 0.41 - 0.71; *P*<.0001), (2) Bleeding Academic Research Consortium 2/3/5 bleeding (RR 0.50, 95% CI 0.26 - 0.94; *P*=.03), (3) adverse cardiovascular events (RR 0.43, 95% CI 0.30 - 0.61; *P*<.00001), and (4) myocardial infarction (RR 0.60, 95% CI 0.41 - 0.89; *P*=.01). However, there were no significant differences between the 2 groups in terms of major adverse cardiovascular and cerebrovascular events, stroke, or cardiovascular mortality.

Conclusions: Compared with aspirin, indobufen demonstrated better safety and was not inferior to aspirin in terms of efficacy, with superior results in some aspects (eg, fewer risks of adverse cardiovascular events and myocardial infarction). Further studies with larger sample sizes or longer follow-up periods may provide additional evidence.

Trial Registration: PROSPERO CRD42024588250; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024588250

(Interact J Med Res 2025;14:e75363) doi:10.2196/75363

KEYWORDS

atherosclerotic disease; antiplatelet therapy; aspirin; indobufen; PRISMA

Introduction

Atherosclerotic diseases include a variety of conditions caused by the buildup of plaque within arteries, leading to their narrowing and hardening [1]. This type of plaque is primarily composed of cholesterol, lipids, and other substances, which progressively restrict blood flow to key organs. Common manifestations include coronary atherosclerosis, intracranial atherosclerosis, and carotid artery and peripheral artery atherosclerotic disease. They increase the risk of cardiovascular events such as heart attacks and strokes. Major risk factors include elevated cholesterol, hypertension, diabetes, smoking, and obesity [2].

The pathogenesis of atherothrombosis is mainly the activation and aggregation of platelets, and platelet reactivity might be associated with the severity of atherosclerosis [3]. Therefore, antiplatelet therapy is the cornerstone of treatment. This activation triggers the release of various signaling molecules, including thromboxane A2 (TXA2), which is synthesized from arachidonic acid via the cyclooxygenase-1 (COX-1) enzyme [4]. TXA2 plays a critical role in amplifying the platelet activation process by promoting the recruitment of additional

XSL•FO
RenderX

platelets to the injury site, thereby increasing platelet aggregation [5]. The importance of the COX-1 enzyme in platelet activation is evident in the mechanism of antiplatelet drugs such as aspirin and indobufen [6,7]. This disruption in the TXA2 pathway significantly lowers the risk of thrombus formation, making COX-1 inhibitors essential for the prevention and treatment of atherosclerotic disease. In summary, the COX-1/TXA2 pathway is central to platelet aggregation, and its inhibition is a key therapeutic strategy for managing atherosclerotic diseases [8].

In patients with atherosclerotic disease, the primary benefit of antiplatelet therapy is the reduction in adverse cardiovascular events related to arterial stenosis, ischemia, and occlusion [9]. The main adverse reaction, however, is mucosal bleeding caused by platelet aggregation inhibition [10]. Given that aspirin has a greater risk of intracranial and gastrointestinal bleeding than other antiplatelet agents do and that some patients may develop resistance, recent studies have focused on alternatives to aspirin [11,12]. Indobufen is gradually attracting the attention of clinicians because of the potential safety profile. Currently, Chinese guidelines acknowledge the feasibility of indobufen as an antiplatelet agent but do not consider it a replacement for aspirin [13]. In contrast, indobufen is less frequently mentioned as a treatment option for patients in the European and American guidelines due to the insufficient clinical evidence. Its further use on a global scale may require additional clinical evidence to attract more attention from clinicians.

In the evaluation of the efficacy and safety of antiplatelet therapy for patients with atherosclerotic disease, commonly assessed outcomes include the incidence of major adverse cardiovascular and cerebrovascular events (MACCEs) [14,15]. MACCEs is a composite end point comprising cardiovascular death, nonfatal myocardial infarction (MI), clinically driven repeated revascularization, definite or probable stent thrombosis, and nonfatal ischemic stroke. Moreover, safety can be evaluated on the basis of the incidence of bleeding events. The Bleeding Academic Research Consortium (BARC) defines the primary safety end point as type 2, 3, or 5 bleeding, also known as major bleeding [16]. Another safety outcome is any bleeding, including mild to severe bleeding.

This paper comprehensively reviewed clinical trials and cohort studies on the efficacy and safety of indobufen in the treatment of atherosclerotic diseases such as stroke and coronary heart disease (CHD), with aspirin as the control. Through meta-analysis, we aimed to integrate existing clinical studies to explore the incidence of MACCEs during the treatment of patients with atherosclerosis with indobufen, evaluate its efficacy, and assess the safety of the drug through the incidence of mucosal bleeding to provide a reference for the clinical application of indobufen and emphasize the necessity of further large-sample prospective studies. Our study aims to contribute to this evidence of indobufen, and we hope that our work will further inform the clinical safety and efficacy of indobufen, inspiring more global clinical research to validate these findings and ultimately contribute to the optimization of pharmacological treatment for atherosclerotic disease.

Methods

Literature Review Design and Literature Search

This study is registered with PROSPERO (CRD42024588250) and was conducted in accordance with the 2020 version of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (the PRISMA checklist is provided in Checklist 1) [17]. Four English databases (Cochrane Library, PubMed, Embase, and Web of Science) and 1 Chinese database (Wanfang) were searched, and the publication deadline for the literature was October 1, 2024. We searched each database to identify MeSH (Medical Subject Headings) (eg, “aspirin,” “indobufen,” and “clinical trial”) and related free terms (eg, Acetylsalicylic Acid*, Acetysal*, Acylpyrin*, Aloxiprimum*, Colfarit*, Ibustrin*, controlled trial*, and Trial*). The search strategy was developed via a combination of MeSH terms and related free terms . Moreover, the references of all included studies were browsed and screened to complement the eligible literature not retrieved. Owing to the limited number of randomized controlled trials (RCTs) in the relevant field, we also included eligible well-conducted cohort studies in our analysis to enhance statistical power and generalizability. This approach of combining RCTs and nonrandomized studies is acceptable when the objective is to evaluate both efficacy and effectiveness in broader clinical settings [18].

Study Selection

For inclusion in the systematic review and meta-analysis, we selected clinical trials that investigated the antiplatelet impact of indobufen and aspirin and their safety in human participants. The following inclusion criteria were applied: (1) the study population comprised adults (aged 18 years and older) with CHD caused by coronary artery atherosclerosis or stroke caused by intracranial atherosclerosis; (2) the intervention was represented by indobufen in the study groups versus aspirin in the control groups; (3) the primary outcome was the incidence of MACCEs or any bleeding or BARC 2/3/5 bleeding; (4) the secondary outcomes were cardiovascular death, MI, and ischemic stroke; adverse cardiovascular events such as coronary thrombus reformation, heart failure, MI, stroke, angina pectoris, and cardiovascular death; stroke; MI; and cardiovascular death; and (5) the studies were randomized clinical trials with either crossover or parallel designs or prospective observational trials. We have supplemented the secondary outcome measures from the registered protocol in order to enhance the guiding value of clinical practice. Our exclusion criteria were as follows: (1) lack of control group studies, (2) no studies reporting cardiovascular or cerebrovascular events or bleeding during the follow-up or research group period, (3) follow-up time of less than 6 months or no mention of follow-up time, and (4) case control studies, cross-sectional studies, ecological studies, case reports, and other descriptive studies. Reviews, conference abstracts, and commentaries were also excluded. Two authors (WHP and LEG) independently searched the appropriate studies from all the databases. Two authors (WHP and LEG) independently screened the studies and removed the papers that were duplicates or not relevant after carefully screening the titles and abstracts. In the next step, we further independently examined the

XSL•FO
RenderX

remaining trials and excluded the ineligible trials according to the full text. We resolved disagreements between the authors by discussion. And a third researcher (HCZ) further resolved any conflicts.

Data Extraction

Two researchers (WHP and LEG) independently extracted the following characteristics from the trials via a standardized data extraction form: year of publication, lead author, country of publication, total sample size of the trial in the intervention and control groups, duration of follow-up (months), age at entry, participants' sex percentage, drug dosage (mg per day), and sample size for the occurrence of outcome indicators.

Statistical Analysis

Our data were managed via Excel, and we used Review Manager 5.3 (The Cochrane Collaboration) and Stata SE 18.0 (Stata Corp LLC) to combine and analyze outcomes from the selected trials. All the data were pooled to calculate the risk ratio (RR) and 95% CIs via a fixed-effects model. Between-trial heterogeneity was assessed by using an I^2 test, and a value of >50% was regarded as considerable heterogeneity. If heterogeneity was high, we used a random-effects model. Statistical significance was defined as a P value <.05. If the number of pooled studies was greater than 10, publication bias was surveyed by visual inspection of the funnel plot and Egger test.

Assessment of Risk of Bias and Sensitivity Analysis

We divided all the included studies into RCTs and non-RCT trials. Two researchers (WHP and LEG) independently assessed the quality of RCTs on the basis of the Cochrane risk-of-bias criteria according to the following criteria [18]: (1) allocation concealment, (2) random sequence generation, (3) blinding of participants and personnel, (4) blinding of outcome assessment, (5) incomplete outcome data, (6) selective reporting, and (7) other probable sources of bias. The Newcastle Ottawa Scale was used to evaluate risk of bias for non-RCTs across domains such as selection, intervention classification and comparability, missing data, outcome measurement, and result reporting [19]. Studies with a score of ≥7 were considered high-quality studies. We performed leave-one-out sensitivity analyses to test the robustness of the results.

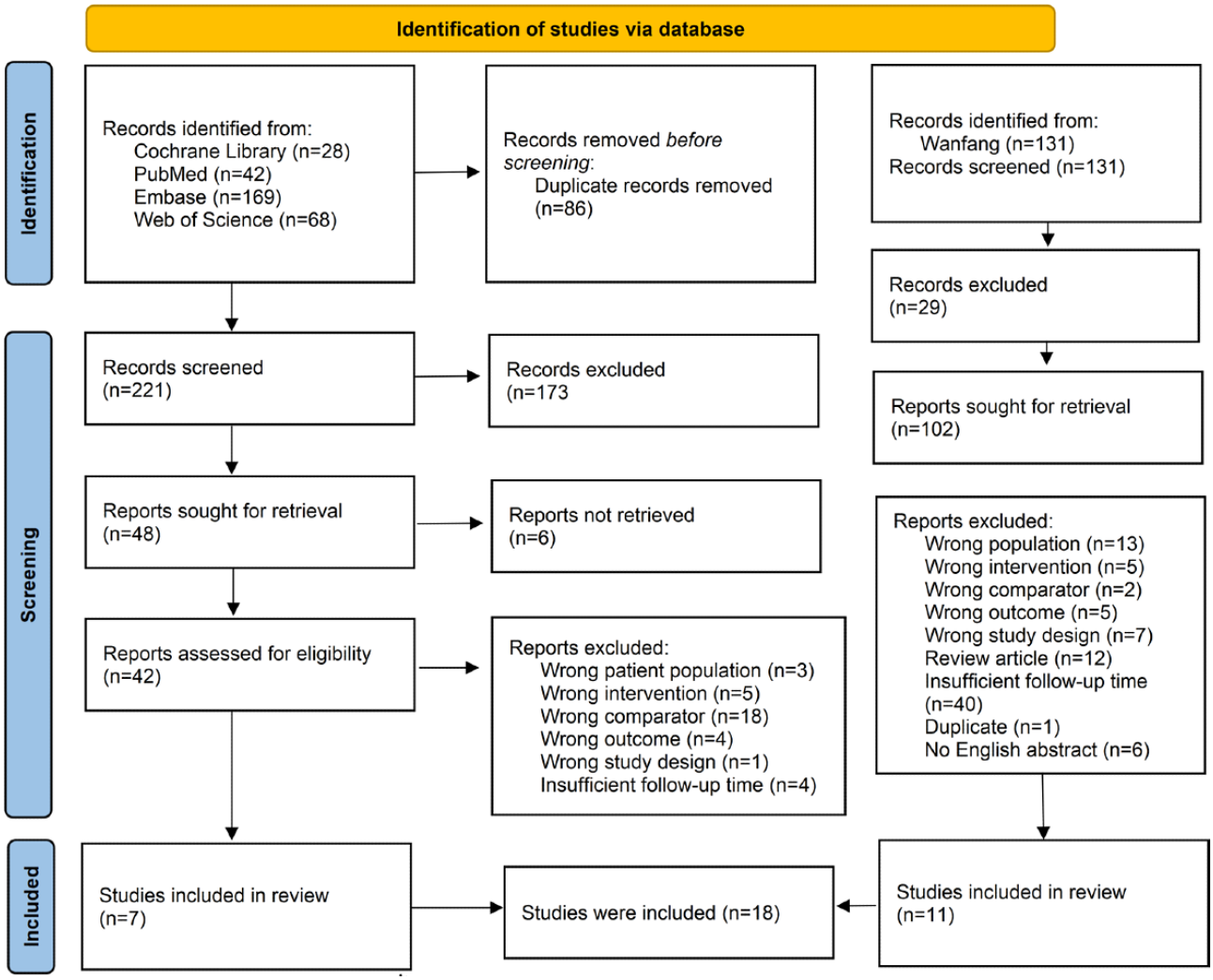
Results

Results of the Literature Search

As reported in Figure 1, 307 papers from the English database were initially retrieved. A total of 86 duplicate studies were removed. Among the remaining 221 papers, 173 were excluded after the titles and abstracts were carefully screened. In the next step, 48 trials were further examined, and 41 additional trials were excluded for the following reasons: the population studied in 3 papers did not meet the inclusion criteria, the intervention methods used in 5 papers did not meet the requirements, 4 studies did not report the required outcome measures, 18 studies did not provide a proper comparison with aspirin, 1 study's design method did not meet the criteria, 4 papers had a follow-up time of less than 6 months or did not mention the follow-up time, and 6 papers did not retrieve the results. Finally, 7 studies (Dai et al [14], Wu et al [15], Bai et al [20], Dai et al [21], Pan et al [22], Shi et al [23], and Ren et al [24]) were included in our meta-analysis. An additional 11 studies (Chen et al [25], Chen [26], Lin et al [27], Liu [28], Ma [29], Tu et al [30], Wen and Cui [31], Wu [32], Yang [33], Rui et al [34], and Zhou et al [35]) from the Chinese database Wanfang were included in the study. A total of 12,981 patients were included in this analysis.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for the search and selection of included studies.



Trial Characteristics

The mean age of the participants varied from 47.20 to 72 years. The sample sizes of the eligible studies ranged from 19 to 2723 participants. All the studies were performed in China. The literature we have included has focused primarily on populations with ischemic stroke caused by cerebral vascular atherosclerosis and CHD, which includes stable angina pectoris, acute coronary syndrome, and coronary artery stent implantation or coronary artery bypass grafting due to CHD. The dose of indobufen given to these participants ranged from 100 mg daily to 100 mg twice a day, and the daily dose of aspirin was 100 mg. The characteristics of the included studies are summarized in Table 1. Figure 2 and Table 2 provide information regarding the quality assessment for the risk of bias for each study included in the meta-analysis.

XSL•FO
RenderX

Table . Characteristics of included studies[a].

Study	Follow-up	Population	Group	Sample size, n	Age (years), mean (SD)	Male, %	Drug and dose	Type
Bai et al [20]	1 year	CABG[b]	Intervention group	76	60.3 (6.6)	75.00	Indobufen 100 mg/day+ Clopidogrel 75 mg/day	RCT[c]
			Control group	76	59.7 (7.2)	81.60	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Chen et al [25]	1 year	CHD[d]	Intervention group	34	66.34 (1.48)	64.70	Indobufen 100 mg twice a day + Clopidogrel 75 mg/day	RCT
			Control group	37	66.39 (1.47)	64.90	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Chen [26]	6 months	CHD	Intervention group	42	48.10 (12.70)	38.10	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	RCT
			Control group	42	47.20 (15.10)	40.48	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Dai et al [14]	1 year	PCI[e]	Intervention group	662	66 (9.4)	72.20	Indobufen 100 mg twice a day + P2Y12 receptor antagonist	Cohort study
			Control group	662	66 (10.5)	70.80	Aspirin 100 mg/day+ P2Y12 receptor antagonist	
Dai et al [21]	1 year	PCI (AMI)[f]	Intervention group	224	72 (11)	62.90	Indobufen 100 mg twice a day + Clopidogrel 75 mg/day	Cohort study
			Control group	224	72 (10)	67.00	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Lin et al [27]	1 year	CCS[g]	Intervention group	37	71.56 (2.19)	54.10	Indobufen 200 mg/day+ Clopidogrel 75 mg/day	Cohort study
			Control group	37	71.66 (2.45)	59.50	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	

XSL•FO
RenderX

Study	Follow-up	Population	Group	Sample size, n	Age (years), mean (SD)	Male, %	Drug and dose	Type
Liu [28]	1 year	PCI	Intervention group	30	58.5 (53.5 - 66.5)	80.00	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	RCT
			Control group	30	60 (56 - 64.25)	73.30	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Ma [29]	6 months	PCI (unstable angina)	Intervention group	41	71.76 (5.85)	68.30	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	RCT
			Control group	41	70.51 (7.55)	65.90	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Pan et al [22]	1 year	Stroke	Intervention group	2715	64.7 (56.3 - 70.9)	64.50	Indobufen 100 mg twice a day+ Placebo	RCT
			Control group	2723	63.8 (56.0 - 70.3)	64.90	Aspirin 100 mg/day+ Placebo	
Rui et al [34]	1 year	ACS[h]	Intervention group	78	68.42 (4.29)	50.00	Indobufen + Clopidogrel	Control study
			Control group	78	68.73 (4.75)	51.30	Aspirin + Clopidogrel	
Shi et al [23]	1 year	PCI (CCS)	Intervention group	52	63.07 (6.46)	96.43	Indobufen 100 mg twice a day	Crossover study
			Control group	52	63.07 (6.46)	96.43	Aspirin 100 mg/day	
Tu et al [30]	6 months	STEMI[i]	Intervention group	53	65.3 (9.8)	56.60	Indobufen 100 mg twice a day+ Ticagrelor 90 mg twice a day	RCT
			Control group	53	64.8 (10.2)	58.49	Aspirin 100 mg/day+ Ticagrelor 90 mg twice a day	
Wen and Cui [31]	6 months	PCI	Intervention group	32	58 (7)	81.25	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	RCT
			Control group	32	57 (9)	75	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	

XSL•FO
RenderX

Study	Follow-up	Population	Group	Sample size, n	Age (years), mean (SD)	Male, %	Drug and dose	Type
WU [32]	1 year	PCI	Intervention group	2258	61.0 (8.3)	67.40	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	RCT
			Control group	2293	61.2 (8.4)	63.10	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Wu et al [15]	6 months	ACS	Intervention group	34	69.88 (4.31)	47.06	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	RCT
			Control group	34	69.68 (4.26)	47.06	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Yang [33]	1 year	ACS	Intervention group	38	70.34 (3.44)	65.70	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	RCT
			Control group	38	71.03 (2.92)	65.70	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Ren et al [24]	18 months	CABG	Intervention group	19	63.84 (7.99)	78.95	Indobufen 100 mg twice a day+ Clopidogrel 75 mg/day	Retrospective cohort study
			Control group	20	64.60 (6.72)	80.00	Aspirin 100 mg/day+ Clopidogrel 75 mg/day	
Zhou et al [35]	1 year	CABG	Intervention group	34	57.63 (2.31)	52.94	Indobufen 0.2 g/day + Clopidogrel 75 mg/day	Retrospective cohort study
			Control group	50	57.54 (2.28)	52	Aspirin 100 mg/day + Clopidogrel 75 mg/day	

[a]Data are presented as means (SDs) and median (IQR).

[b]CABG: coronary artery bypass grafting.

[c] RCT: randomized controlled trial.

[d]CHD: coronary heart disease.

[e]PCI: percutaneous coronary intervention.

[f]AMI: acute myocardial infarction.

[g]CCS: chronic coronary syndromes.

[h]ACS: acute coronary syndromes.

[i]STEMI: ST-segment elevation myocardial infarction.

XSL•FO
RenderX

Figure 2. Risk of bias summary: review authors' judgments about each risk of bias item for RCTs [15,20,22,25,26,28,29,30,31,32,33,34].

Table . Quality assessment of non–randomized controlled trial according to Newcastle Ottawa Scale.

Study	Selection				Comparability		Results			Total score
	S1	S2	S3	S4	C1	C2	R1	R2	R3	
Dai et al [14]	P[a]	P	P	P	P	N[b]	P	P	P	8
Dai et al [21]	P	P	P	P	P	N	P	P	P	8
Lin et al [27]	P	P	P	P	P	N	P	P	N	7
Shi et al [23]	P	P	P	P	P	N	P	P	N	7
Ren et al [24]	P	P	P	P	P	N	P	P	P	8
Zhou et al [35]	P	P	P	P	P	N	P	P	N	7

[a]P: 1 score.

[b]N: zero score.

XSL•FO
RenderX

Meta-Analysis Results

Major Adverse Cardiovascular and Cerebrovascular Events

Five studies reported MACCEs (Dai et al [14], Wu et al [15], Bai et al [20], Dai et al [21], and Pan et al [22]). The pooled results revealed no significant difference in MACCEs with indobufen compared with aspirin (RR 1.12, 95% CI 0.98 - 1.28; I^2=0%; *P*=.09; Figure S1 in Multimedia Appendix 1 [14,15,20-22]).

Any Bleeding

Seventeen studies reported any bleeding (Dai et al [14], Wu et al [15], Bai et al [20], Dai et al [21], Pan et al [22], Shi et al [23], Ren et al [24], Chen et al [25], Chen [26], Lin et al [27], Liu [28], Ma [29], Tu et al [30], Wen and Cui [31], Yang [33], Rui et al [34], and Zhou et al [35]). However, 1 study was excluded because there were no outcome events in either the experimental group or the control group. The combined results revealed a significant reduction in any bleeding with indobufen compared with aspirin (RR 0.54, 95% CI 0.41 - 0.71; I^2=51%; *P*<.0001; Figure S2 in Multimedia Appendix 1 [14,15-21,22-29,30-33,34,35]).

BARC 2/3/5 Bleeding

Five studies reported BARC type 2, 3, or 5 bleeding (BARC 2/3/5 bleeding; Dai et al [14], Wu et al [15], Dai et al [21], Shi et al [23], and Liu [28]). Two studies were excluded from the pooled meta-analysis because both groups in these studies reported zero events. The heterogeneity (I^2=84%; *P*=.002) was high, so we used a random-effects model to assess the outcome. The combined results revealed a reduction in BARC 2/3/5 bleeding with indobufen compared with aspirin (RR 0.50, 95% CI 0.26 - 0.94; I^2=84%; *P*=.03; Figure S3 in Multimedia Appendix 1 [14,15,21,23,28]).

Adverse Cardiovascular Events

Thirteen studies reported adverse cardiovascular events (Bai et al [20], Ren et al [24], Chen et al [25], Chen [26], Lin et al [27], Liu [28], Ma [29], Tu et al [30], Wen and Cui [31], Wu [32], Yang [33], Rui et al [34], and Zhou et al [35]). The pooled results revealed a significant decrease in adverse cardiovascular events with indobufen compared with aspirin (RR 0.43, 95% CI 0.30 - 0.61; I^2=0%; *P*<.00001; Figure S4 in Multimedia Appendix 1 [20,24-35]).

Stroke

Seventeen studies reported stroke (Dai et al [14], Wu et al [15], Bai et al [20], Dai et al [21], Pan et al [22], Ren et al [24], Chen et al [25], Chen [26], Lin et al [27], Liu [28], Ma [29], Tu et al [30], Wen and Cui [31], Wu [32], Yang [33], Rui et al [34], and Zhou et al [35]). Five studies were excluded because of zero events. The pooled results revealed no significant difference in the incidence of stroke between patients treated with indobufen and those treated with aspirin (RR 1.07, 95% CI 0.92 - 1.25; I^2=0%; *P*=.37; Figure S5 in Multimedia Appendix 1 [14,15,20-22,24-35]).

Myocardial Infarction

Seventeen studies reported MI (Dai et al [14], Wu et al [15], Bai et al [20], Dai et al [21], Pan et al [22], Ren et al [24], Chen et al [25], Chen [26], Lin et al [27], Liu [28], Ma [29], Tu et al [30], Wen and Cui [31], Wu [32], Yang [33], Rui et al [34], and Zhou et al [35]). Four studies were excluded because of zero events. The pooled results revealed a significant decrease in MI with indobufen compared with aspirin (RR 0.60, 95% CI 0.41 - 0.89; I^2=0%; *P*=.01; Figure S6 in Multimedia Appendix 1 [14,15,20-22,24-35]).

Cardiovascular Death

Sixteen studies reported cardiovascular death (Dai et al [14], Wu et al [15], Bai et al [20], Dai et al [21], Pan et al [22], Chen et al [25], Chen [26], Lin et al [27], Liu [28], Ma [29], Tu et al [30], Wen and Cui [31], Wu [32], Yang [33], Rui et al [34], and Zhou et al [35]). Eight studies were excluded because of zero events. The pooled results revealed no significant difference in cardiovascular death between indobufen and aspirin (RR 1.13, 95% CI 0.71 - 1.79; I^2=0%; *P*=.61; Figure S7 in Multimedia Appendix 1 [14,15,20-22,25-35]).

Sensitivity Analysis

In this meta-analysis, the results of any bleeding and BARC studies revealed a high level of heterogeneity, with I^2 values of 51% and 84%, indicating substantial heterogeneity, possibly due to the small sample size. To further explore the source of this heterogeneity, we conducted a sensitivity analysis. This involved sequentially excluding individual studies to assess their influence on the overall pooled estimate. Although the limited number of studies restricts the depth of the analysis, the results showed that the pooled effect estimate remained stable, indicating relative robustness (Figures 3 and 4). However, variations in heterogeneity were observed when individual studies were excluded, suggesting that 1 or more studies might have contributed to the observed heterogeneity. We attribute the source of the heterogeneity primarily to the use of the second antiplatelet agent in the dual antiplatelet therapy (DAPT) regimen, as well as to variations in sample size and concomitant use of proton pump inhibitors.

XSL•FO
RenderX

Figure 3. Sensitivity analysis of any bleeding [14,15-21,22-30,31-34,35].

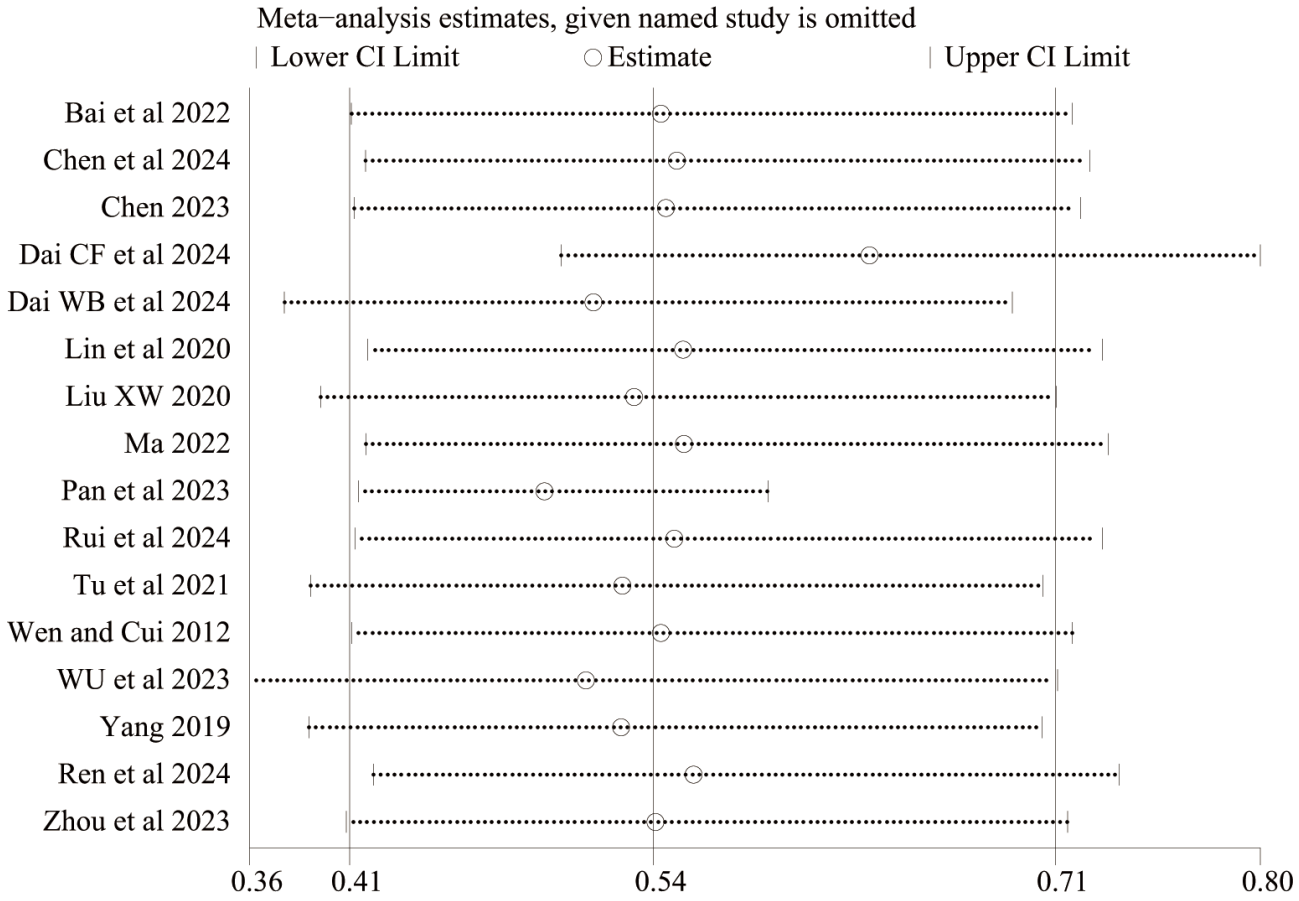
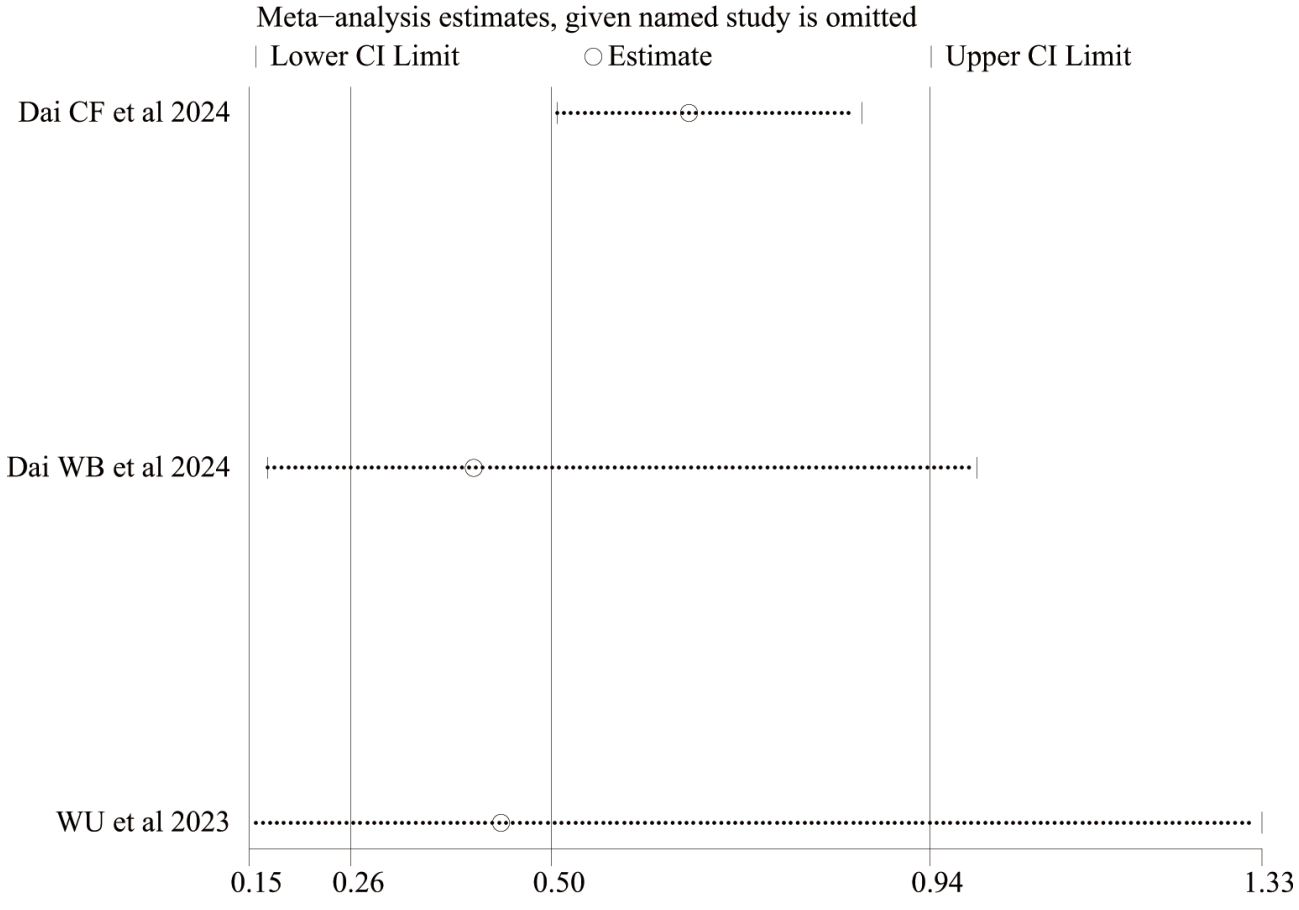


Figure 4. Sensitivity analysis of BARC 2/3/5 bleeding [14,15,21].



XSL•FO
RenderX

Publication Bias

For outcomes with more than 10 papers included in the literature, we plotted funnel plots and performed Egger tests. There are 16, 13, 12, and 13 studies on outcomes of any bleeding events, adverse cardiovascular events, stroke events, and MI, respectively. The *P* values of the Egger tests for any bleeding or adverse cardiovascular events were .249 and .235, respectively, both of which are greater than .05. We believe that the publication bias is small, but the *P* values of the Egger tests for stroke and MI were .000 and .001, indicating a certain degree of publication bias. All funnel plots are included in Multimedia Appendix 1.

Discussion

Principal Findings

In this systematic review and meta-analysis involving 18 clinical trials and 12,981 patients, we compared antiplatelet therapy involving indobufen with that involving aspirin. Although we did not impose any language restrictions for the included studies, all the clinical trials ultimately retrieved were conducted in China. This may be because indobufen is not widely approved as a routine antiplatelet agent in Europe and the Americas and is used only as a reference drug in specific studies [36]. These results suggest that, compared with aspirin, indobufen can effectively reduce the incidence of mucosal bleeding (including both major and minor bleeding) in patients with cardiovascular and cerebrovascular atherosclerosis. Additionally, on the basis of the studies included in our analysis, indobufen was shown to reduce the incidence of adverse cardiovascular events (a composite outcome of coronary thrombus reformation, heart failure, MI, stroke, angina pectoris, and cardiovascular death) as well as MI. However, MACCEs, stroke, and cardiovascular death rates were similar between the 2 groups, which may be related to the limited follow-up time. The heterogeneity for any bleeding and BARC 2/3/5 bleeding was high, but the sensitivity analysis revealed no difference in outcomes. There was significant publication bias for stroke, MI, and cardiovascular death. The effects of indobufen on MACCEs, stroke, and cardiovascular death were similar to those of aspirin, demonstrating no inferiority in efficacy.

Antiplatelet therapy in patients with atherosclerosis remains a topic of considerable debate. Current guidelines recommend aspirin for secondary prevention in patients with coronary atherosclerotic heart disease, and DAPT is advised for patients with a high risk of thrombosis and a low risk of bleeding or those undergoing interventional treatment [37]. For patients who require DAPT but are intolerant to aspirin, indobufen is being investigated as a potential alternative to aspirin [38,39].

The antiplatelet effect of aspirin is achieved by irreversibly inhibiting and permanently blocking COX-1 in platelets, thereby preventing the formation of thromboxane A2 (TXA2) [40]. This effect lasts for the lifespan of the platelet (7 - 10 days) [41]. In contrast, indobufen is a reversible COX-1 inhibitor, meaning that it temporarily inhibits the enzyme but does not cause persistent inhibition [42]. This reversible action allows for shorter-term effects on platelet function than does aspirin [43]. The inhibitory effects of indobufen on platelet function returned to baseline within 24 hours after discontinuation. Additionally, indobufen does not substantially impact prostacyclin, potentially minimizing the risk of aspirin-gastric side effects [25]. While both drugs seek to reduce platelet aggregation and associated thrombosis, the irreversible inhibition of aspirin leads to protracted platelet inhibition, whereas indobufen offers a reversible and potentially safer alternative, particularly in patients at greater risk for bleeding.

Aspirin's first-line status in the prevention and treatment of cardiovascular and cerebrovascular diseases is supported by a large body of evidence-based research [10,44,45]. In comparison, although the clinical application of indobufen has evolved, it remains very limited at present. Initially, owing to its reversible inhibition of COX, it was not recommended, as it was considered insufficient to effectively inhibit platelet aggregation in vivo [46]. However, subsequent clinical trials have led some meta-analyses to consider indobufen similarly to aspirin. Cataldo G et al [47], in their combined analysis of 2 multicenter coronary artery bypass grafting studies, concluded that indobufen is as effective as aspirin and dipyridamole in preventing graft occlusion. Bhana and McClellan [48], in their review, suggested that indobufen may be an effective alternative to aspirin for patients with nonrheumatic atrial fibrillation who have contraindications to anticoagulation or a greater risk of bleeding. In a meta-analysis of 9 clinical studies, Zhang et al [49] reported that indobufen had fewer gastrointestinal reactions, a lower risk of bleeding, and could be a viable option for patients at high risk of bleeding or with gastrointestinal ulcers.

Strengths and Limitations

The strengths of our study include the following: compared with previous studies, our meta-analysis incorporated a larger number of clinical trials and patients with homogeneous outcome definitions and a large patient population, making the results more generalizable to the broader population. Additionally, the follow-up duration in all the included studies exceeded 6 months, enhancing the credibility of our findings. We also conducted publication bias and sensitivity analyses to further emphasize the reliability of this meta-analysis.

There are several limitations to this study. All the studies included focused on Chinese populations, which may limit the generalizability of our findings to other regions of the world. But we believe that with the accumulation of further clinical evidence in China, the safety profile of indobufen as a potential alternative to aspirin will gain stronger support, paving the way for the development of larger multicenter studies with longer follow-up periods and more diverse populations across different regions—including global clinical trials. Additionally, because antiplatelet therapy is a long-term treatment requiring good patient adherence, the included trials were open-label, which might raise concerns about potential bias, although none of the studies carried a high risk of bias. Owing to the lack of available data, we were unable to perform subgroup analyses on the basis of individual patient characteristics, and the efficacy and safety of indobufen in specific populations require further investigation.

XSL•FO
RenderX

Conclusions

Our meta-analysis indicates that, compared with aspirin, indobufen offers a safer profile for antiplatelet therapy in patients with cardiovascular and cerebrovascular atherosclerosis, as it reduces the incidence of mucosal bleeding. In terms of efficacy, indobufen is associated with a lower incidence of adverse cardiovascular events and MI, while its impact on MACCEs and cardiovascular mortality is similar to that of aspirin. Indobufen demonstrated no overall inferiority in effectiveness, with superior results in some aspects (adverse cardiovascular events and MI). This provides additional robust evidence supporting the feasibility of using indobufen as an alternative antiplatelet therapy, particularly in patients who are intolerant to aspirin.

Acknowledgments

This study is supported by major projects of the National Social Science Fund (18ZDA086-4).

Data Availability

The original contributions presented in the study are included in the paper or supporting information material, and further inquiries can be directed to the corresponding authors.

Authors' Contributions

HZ designed the systematic review plan; WP and LG conducted research screening, as well as data extraction and statistical analysis; and WP wrote the manuscript. All authors contributed to the critical revision of the manuscript draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Forest plots for the outcomes. Figure S1: Major adverse cardiovascular and cerebrovascular events (risk ratio 1.12, CI 0.98-1.28; *P*=.09); Figure S2: any bleeding events (risk ratio 0.54, CI 0.41-0.71; *P*<.0001); Figure S3: Bleeding Academic Research Consortium 2/3/5/5 bleeding (risk ratio 0.50, CI 0.26-0.94; *P*=.03); Figure S4: adverse cardiovascular events (risk ratio 0.43, CI 0.30-0.61; *P*<.00001); Figure S5: stroke events (risk ratio 1.07, CI 0.92-1.25; *P*=.37); Figure S6: myocardial infarction events (risk ratio 0.60, CI 0.41-0.89; *P*=.01); and Figure S7: cardiovascular death associated with the use of indobufen and aspirin (risk ratio 1.13, CI 0.71-1.79; *P*=.61). RR: risk ratio.

[PDF File, 720 KB - i-jmr_v14i1e75363_app1.pdf]

Checklist 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.

[DOCX File, 30 KB - i-jmr_v14i1e75363_app2.docx]

References

1. Fuster V, Badimon L, Badimon JJ, Chesebro JH. The pathogenesis of coronary artery disease and the acute coronary syndromes (2). N Engl J Med 1992 Jan 30;326(5):310-318. [doi: 10.1056/NEJM199201303260506] [Medline: 1728735]
2. Pasternak RC, Criqui MH, Benjamin EJ, et al. Atherosclerotic vascular disease conference. Circulation 2004 Jun;109(21):2605-2612. [doi: 10.1161/01.CIR.0000128518.26834.93]
3. Yun KH, Mintz GS, Witzenbichler B, et al. Relationship between platelet reactivity and culprit lesion morphology: an assessment from the ADAPT-DES intravascular ultrasound substudy. JACC Cardiovasc Imaging 2016 Jul;9(7):849-854. [doi: 10.1016/j.jcmg.2015.08.019] [Medline: 26897675]
4. Kirkby NS, Lundberg MH, Harrington LS, et al. Cyclooxygenase-1, not cyclooxygenase-2, is responsible for physiological production of prostacyclin in the cardiovascular system. Proc Natl Acad Sci U S A 2012 Oct 23;109(43):17597-17602. [doi: 10.1073/pnas.1209192109] [Medline: 23045674]
5. Angiolillo DJ. The evolution of antiplatelet therapy in the treatment of acute coronary syndromes: from aspirin to the present day. Drugs (Abingdon Engl) 2012 Nov 12;72(16):2087-2116. [doi: 10.2165/11640880-000000000-00000] [Medline: 23083110]
6. Smith JB, Araki H, Lefer AM. Thromboxane A2, prostacyclin and aspirin: effects on vascular tone and platelet aggregation. Circulation 1980 Dec;62(6 Pt 2):V19-V25. [Medline: 7002350]
7. Liu X, Lv X, Peng Y, et al. Clopidogrel with indobufen or aspirin in minor ischemic stroke or high-risk transient ischemic attack: a randomized controlled clinical study. BMC Neurol 2024 Mar 1;24(1):81. [doi: 10.1186/s12883-024-03585-4] [Medline: 38429754]

XSL•FO
RenderX

8. Bunimov N, Laneuville O. Cyclooxygenase inhibitors: instrumental drugs to understand cardiovascular homeostasis and arterial thrombosis. Cardiovasc Hematol Disord Drug Targets 2008 Dec;8(4):268-277. [doi: 10.2174/187152908786786250] [Medline: 19075637]
9. Hennekens CH. Update on aspirin in the treatment and prevention of cardiovascular disease. Am J Manag Care 2002 Dec;8(22 Suppl):S691-S700. [Medline: 12512736]
10. Zusman RM, Chesebro JH, Comerota A, et al. Antiplatelet therapy in the prevention of ischemic vascular events: literature review and evidence-based guidelines for drug selection. Clin Cardiol 1999 Sep;22(9):559-573. [doi: 10.1002/clc.4960220905] [Medline: 10486695]
11. US Preventive Services Task Force, Davidson KW, Barry MJ, et al. Aspirin use to prevent cardiovascular disease: us preventive services task force recommendation statement. JAMA 2022 Apr 26;327(16):1577-1584. [doi: 10.1001/jama.2022.4983] [Medline: 35471505]
12. Horiuchi H. Recent advance in antiplatelet therapy: the mechanisms, evidence and approach to the problems. Ann Med 2006;38(3):162-172. [doi: 10.1080/07853890600640657] [Medline: 16720431]
13. Chinese Society of Cardiology, Chinese Medical Association, Editorial Board of Chinese Journal of Cardiology. Chinese guidelines for the diagnosis and management of patients with chronic coronary syndrome. Zhonghua Xin Xue Guan Bing Za Zhi 2024;52(6):589-614. [doi: 10.3760/cma.j.cn112148-20240325-00168]
14. Dai C, Liu M, Yang Z, et al. Real-world performance of indobufen versus aspirin after percutaneous coronary intervention: insights from the ASPIRATION registry. BMC Med 2024 Apr 2;22(1):148. [doi: 10.1186/s12916-024-03374-3] [Medline: 38561738]
15. Wu H, Xu L, Zhao X, et al. Indobufen or aspirin on top of clopidogrel after coronary drug-eluting stent implantation (OPTION): a randomized, open-label, end point–blinded, noninferiority trial. Circulation 2023 Jan 17;147(3):212-222. [doi: 10.1161/CIRCULATIONAHA.122.062762] [Medline: 36335890]
16. Mehran R, Rao SV, Bhatt DL, et al. Standardized bleeding definitions for cardiovascular clinical trials: a consensus report from the Bleeding Academic Research Consortium. Circulation 2011 Jun 14;123(23):2736-2747. [doi: 10.1161/CIRCULATIONAHA.110.009449] [Medline: 21670242]
17. 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]
18. Higgins JPT, et al. Cochrane Handbook for Systematic Reviews of Interventions Version 65: Cochrane; 2024. URL: www.training.cochrane.org/handbook [accessed 2025-07-06]
19. Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hosp Res Inst 2011;2(1):1-12 [FREE Full text]
20. Bai C, Li JX, Yu Y, et al. A randomized controlled trial of indobufen versus aspirin in the prevention of bridging restenosis after coronary artery bypass grafting. Zhonghua Xin Xue Guan Bing Za Zhi 2022 May 24;50(5):466-470. [doi: 10.3760/cma.j.cn112148-20210701-00560] [Medline: 35589595]
21. Dai WB, Ren JY, Hu ST, et al. The safety and efficacy of indobufen or aspirin combined with clopidogrel in patients with acute myocardial infarction after percutaneous coronary intervention. Platelets 2024 Dec;35(1):2364748. [doi: 10.1080/09537104.2024.2364748] [Medline: 39115322]
22. Pan Y, Meng X, Yuan B, et al. Indobufen versus aspirin in patients with acute ischaemic stroke in China (INSURE): a randomised, double-blind, double-dummy, active control, non-inferiority trial. Lancet Neurol 2023 Jun;22(6):485-493. [doi: 10.1016/S1474-4422(23)00113-8] [Medline: 37121237]
23. Shi QP, Luo XY, Zhang B, et al. Effect of indobufen vs. aspirin on platelet accumulation in patients with stable coronary heart disease after percutaneous coronary intervention: an open-label crossover study. Front Pharmacol 2022;13:950719. [doi: 10.3389/fphar.2022.950719] [Medline: 36052139]
24. Ren Y, Zhu Y, Yan Q, Jin H, Luo H. Multicenter retrospective cohort study demonstrates superior safety profile of indobufen over aspirin for Post-CABG antiplatelet therapy. Front Pharmacol 2024;15:1474150. [doi: 10.3389/fphar.2024.1474150] [Medline: 39403145]
25. Chen B, Yang YJ, Song JJ. Observation on the effect of indobufen in the treatment of coronary heart disease combined with non valvular atrial fibrillation. Technol Health 2024;3(11):45-48 [FREE Full text]
26. Chen YC. The effect of indobufen combined with clopidogrel on myocardial perfusion, NT-proBNP, and MACE in patients with coronary heart disease after PCI. Clin Res 2023;31(10):53-56. [doi: 10.12385/j.issn.2096-1278(2023)10-0053-04]
27. Lin F, Huang DS, Fu MW. The effect of clopidogrel and indobufen dual antiplatelet therapy on patients with stable coronary heart disease after coronary stenting. China Prac Med 2020;15(31):95-97. [doi: 10.14163/j.cnki.11-5547/r.2020.31.040]
28. Liu XW. Indobufen versus aspirin combined with clopidogrel used for dual antiplatelet therapy after drug-eluting stent implantation. 2020 URL: https://med.wanfangdata.com.cn/Paper/Detail?id=DegreePaper_D02338473&dbid=WF_XW [accessed 2025-07-29]
29. Ma K. Efficacy and safety study of indobufen combined with clopidogrel in patients with unstable angina pectoris after PCI. : Hebei Medical University; 2022.

XSL•FO
RenderX

30. Tu ZT, Pei F, Yu CQ. Effect of ticagrelor combined with indobufen on clinical efficacy and prognosis of patients with acute ST-segment elevation myocardial infarction. Chongqing Med 2021;50(21):3660-3663. [doi: 10.3969/j.issn.1671-8348.2021.21.013]
31. Wen JL, Cui XY. To explore the prevention effect of indobufen on PCI patients with coronary artery restenosis. J TCM Univ Hunan 2012;32(10):29-30. [doi: 10.3969/j.issn.1674-070X.2012.10.013.029.03]
32. WU YQ. Clinical comparison of indobufen and aspirin combined with clopidogrel in the treatment of acute coronary syndrome. Chin Foreign Med Res 2020;18(32):40-42. [doi: 10.14033/j.cnki.cfmr.2020.32.015]
33. Yang S. Efficacy and safety of indobufen plus clopidogrel for postoperative PCI in patients with ACS who are at high risk for gastrointestinal bleeding. : Hebei Medical University; 2019.
34. Rui SH, Wang CF, Li YH. Application of indobuprofen combined with clopidogrel in patients with ACS after PCI. Chin J Health Care Med 2024;26(3):270-273. [doi: 10.3969/j.issn.1674-3245.2024.03.004]
35. Zhou CY, Su GB, Liu XC. Clinical efficacy and safety of indobufen combined with clopidogrel in preventing vascular restenosis in saphenous vein grafts following coronary artery bypass grafting. Chin J Ration Drug Use 2023;20(5):122-126. [doi: 10.3969/j.issn.2096-3327.2023.05.020]
36. Bianco M, Biolè C, Spirito A, et al. Management of ST-elevation myocardial infarction in a patient with aspirin hypersensitivity without oral P2Y12 inhibitors. G Ital Cardiol (Rome) 2022 Jan;23(1):75-77. [doi: 10.1714/3715.37065] [Medline: 34985465]
37. Writing Committee Members, Virani SS, Newby LK, et al. 2023 AHA/ACC/ACCP/ASPC/NLA/PCNA guideline for the management of patients with chronic coronary disease: a report of the American Heart Association/American College of Cardiology Joint Committee on Clinical Practice Guidelines. J Am Coll Cardiol 2023 Aug 29;82(9):833-955. [doi: 10.1016/j.jacc.2023.04.003] [Medline: 37480922]
38. Wiseman LR, Fitton A, Buckley MM. Indobufen. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy in cerebral, peripheral and coronary vascular disease. Drugs (Abingdon Engl) 1992 Sep;44(3):445-464. [doi: 10.2165/00003495-199244030-00009] [Medline: 1382938]
39. Latib A, Ielasi A, Ferri L, et al. Aspirin intolerance and the need for dual antiplatelet therapy after stent implantation: a proposed alternative regimen. Int J Cardiol 2013 May 25;165(3):444-447. [doi: 10.1016/j.ijcard.2011.08.080] [Medline: 21968077]
40. Catella-Lawson F, Reilly MP, Kapoor SC, et al. Cyclooxygenase inhibitors and the antiplatelet effects of aspirin. N Engl J Med 2001 Dec 20;345(25):1809-1817. [doi: 10.1056/NEJMoa003199] [Medline: 11752357]
41. Burch JW, Stanford N, Majerus PW. Inhibition of platelet prostaglandin synthetase by oral aspirin. J Clin Invest 1978 Feb;61(2):314-319. [doi: 10.1172/JCI108941] [Medline: 413839]
42. Liu J, Sun P, Qi X. Reversible and non-competitive inhibition of cyclooxygenase by indobufen for efficient antiplatelet action and relief of gastrointestinal irritation. Pharmaceutics 2023 Aug 14;15(8):2135. [doi: 10.3390/pharmaceutics15082135] [Medline: 37631348]
43. Tamassia V, Corvi G, Fuccella LM, Moro E, Tosolini G, Tremoli E. Indobufen (K 3920), a new inhibitor of platelet aggregation: effect of food on bioavailability, pharmacokinetic and pharmacodynamic study during repeated oral administration to man. Eur J Clin Pharmacol 1979 Jun 12;15(5):329-333. [doi: 10.1007/BF00558436] [Medline: 456404]
44. Gaziano JM, Brotons C, Coppolecchia R, et al. Use of aspirin to reduce risk of initial vascular events in patients at moderate risk of cardiovascular disease (ARRIVE): a randomised, double-blind, placebo-controlled trial. Lancet 2018 Sep 22;392(10152):1036-1046. [doi: 10.1016/S0140-6736(18)31924-X] [Medline: 30158069]
45. Johnston SC, Amarenco P, Albers GW, et al. Ticagrelor versus aspirin in acute stroke or transient ischemic attack. N Engl J Med 2016 Jul 7;375(1):35-43. [doi: 10.1056/NEJMoa1603060] [Medline: 27160892]
46. Yang M, Ye Z, Mei L, et al. Pharmacodynamic effects of indobufen compared with aspirin in patients with coronary atherosclerosis. Eur J Clin Pharmacol 2021 Dec;77(12):1815-1823. [doi: 10.1007/s00228-021-03177-y] [Medline: 34331551]
47. Cataldo G, Heiman F, Lavezzari M, Marubini E. Indobufen compared with aspirin and dipyridamole on graft patency after coronary artery bypass surgery: results of a combined analysis. Coron Artery Dis 1998;9(4):217-222. [doi: 10.1097/00019501-199809040-00007] [Medline: 9649928]
48. Bhana N, McClellan KJ. Indobufen: an updated review of its use in the management of atherothrombosis. Drugs Aging 2001;18(5):369-388. [doi: 10.2165/00002512-200118050-00007] [Medline: 11392445]
49. Zhang X, Yan Q, Jiang J, Luo H, Ren Y. Safety and efficacy of aspirin and indobufen in the treatment of coronary heart disease: a systematic review and meta-analysis. Front Cardiovasc Med 2024;11:1412944. [doi: 10.3389/fcvm.2024.1412944] [Medline: 39211768]

Abbreviations

BARC: Bleeding Academic Research Consortium
CHD: coronary heart disease
COX-1: cyclooxygenase-1
DAPT: dual antiplatelet therapy

XSL•FO
RenderX

MACCE: major adverse cardiovascular and cerebrovascular event
MeSH: Medical Subject Headings
MI: myocardial infarction
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: risk ratio
TXA2: thromboxane A2

Edited by TDA Cardoso; submitted 01.04.25; peer-reviewed by A Adekola, A Rawat, U Modebelu; revised version received 01.06.25; accepted 05.06.25; published 20.08.25.

Please cite as:
Pan W, Guan L, Zhang H
Safety and Efficacy of Aspirin and Indobufen in the Treatment of Atherosclerotic Diseases: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e75363
URL: https://www.i-jmr.org/2025/1/e75363
doi:10.2196/75363

©Wenhao Pan, Linger Guan, Haicheng Zhang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 20.8.2025. 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

Reasons for Crown Failures in Primary Teeth: Systematic Review and Meta-Analysis

Stephan Lampl[1], MClinDent; Deepa Gurunathan[1], MDS; Deepak Mehta[2,3], PhD; Krithikadatta Jogikalmat[2], MDS

[1]Department of Pediatric Dentistry, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, 162, Poonamallee High Rd, Velappanchavadi, Chennai, India

[2]Department of Cariology, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, Chennai, India

[3]Department of Operative Dentistry, Tohoku Graduate School of Dentistry, Tohoku University, Sendai, Japan

Corresponding Author:
Deepa Gurunathan, MDS
Department of Pediatric Dentistry, Saveetha Dental College and Hospitals, Saveetha Institute of Medical and Technical Sciences, 162, Poonamallee High Rd, Velappanchavadi, Chennai, India

Abstract

Background: Understanding long-term retention rates and complications associated with different materials for fabricating pediatric crowns for primary teeth is crucial for material selection and optimizing clinical outcomes.

Objectives: This systematic review aimed to descriptively analyze the crown-retention rates and complications associated with crown retention, as well as the biological and technical complications of pediatric crowns, for primary teeth. The meta-analysis reported herein was performed to estimate long-term (3-year and 5-year) retention rates of these pediatric crowns fabricated using various materials.

Methods: Using the PICOS (Population, Intervention, Comparison, Outcomes, and Study design) paradigm, a systematic search was conducted between July and August 2023 in the Cochrane, Embase, and PubMed databases to identify randomized controlled trials (RCTs) and clinical (prospective and retrospective) studies reporting retention rates, complications of crown retention, and biological and technical complications. After selecting studies with a predefined set of selection criteria, data from included studies were used for a systematic review aimed at a descriptive analysis of factors associated with the failure of crowns for primary teeth. Data from the included RCTs were used for meta-analysis, wherein 3-year and 5-year crown-retention rates were estimated using Poisson regression models.

Results: This systematic review included 13 RCTs and 5 clinical studies on dental crowns for primary teeth, involving 454 children (1172 crowns) in RCTs and 810 children (2667 crowns) in clinical studies. The median follow-up durations were 12 months for RCTs and 20.8 months for clinical studies, with a 10.6% (124/1172) dropout rate in RCTs. Meta-analysis of pooled 5-year retention rates for different crown materials revealed the following retention rates: 88.90% for compomer crowns, 92.18% for composite resin crowns, 90.30% for resin-modified glass ionomer cement (RMGIC) crowns, and 97.88% for stainless steel crowns. Additionally, strip crowns exhibited a retention rate of 83.48%, while zirconia crowns had a retention rate of 97.09%. Poisson regression estimated 3-year and 5-year crown-retention rates, indicating good outcomes across materials. Complications included secondary caries (up to 21.8% in zirconia crowns) and marginal adaptation issues (up to 22.2% in compomer crowns). These findings highlight material-specific considerations necessary for optimizing outcomes in pediatric dental crown treatments.

Conclusion: While retentive complications such as chipping, material loss, and fractures do occur across materials, compomer, composite resin, stainless steel, strip, and zirconia crowns all have clinically acceptable retention rates. However, the differences in biological and technical complications between materials may provide insights for selecting appropriate materials for pediatric crowns based on clinical considerations.

Trial Registration: PROSPERO CRD42023442266; https://www.crd.york.ac.uk/PROSPERO/view/CRD42023442266

International Registered Report Identifier (IRRID): RR2-10.2196/51505

(Interact J Med Res 2025;14:e57958) doi:10.2196/57958

KEYWORDS

primary teeth; pediatrics; children; biological complications; technical complications; survival rates; dental; oral; dentist; synthesis; review methods; search; systematic; meta-analysis; complication; crown

XSL•FO
RenderX

Introduction

Dental caries is a highly prevalent chronic disease affecting millions of children worldwide, leading to pain, infection, and difficulties in eating and speaking [1-3]. Untreated caries of the primary teeth may affect over 621 million children globally [1]. Early intervention with fillings is essential to manage caries, but in cases where the tooth structure is significantly compromised, pediatric dental crowns become necessary to preserve the tooth's functionality and prevent premature loss [4].

Several studies have documented the usefulness and efficacy of materials such as stainless steel, zirconia, composite resin, and polycarbonate materials for fabricating pediatric dental crowns [5-8]. Despite notable advancements in dental materials and restorative techniques, various biological and technical factors persist in influencing the longevity and acceptability of pediatric dental crowns [8]. Biological factors such as the oral microbiome, the presence of cariogenic bacteria, and the immune response play significant roles in the success or failure of pediatric dental crowns [8]. Additionally, technical factors including crown preparation, cementation methods, and occlusal adjustment can impact the longevity and performance of these restorations [8]. Understanding these factors is crucial for clinicians to make informed decisions regarding treatment planning and material selection.

Chisini et al [8], in their systematic review, highlighted that composite resin crowns exhibited the lowest annual failure rate (1.7% to 12.9%), while stainless steel crowns had the highest success rate (96.1%). This report [8], published in 2018, identified secondary caries as the main reason for the failure of pediatric crowns and recommended an anticariogenic, health-promoting approach. Furthermore, several systematic reviews (published between 2021 and 2023) have also reported on the efficacy of individual restorative materials such as stainless steel, zirconia, and prefabricated crowns [9-11]. Since the publication of these systematic reviews, several randomized controlled trials (RCTs) have been published, warranting an updated systematic review to include these additional articles.

In the context of overall public health, ensuring the longevity and success of pediatric dental crowns not only addresses the immediate dental needs of children but also contributes to their overall health and well-being [12,13]. Improving the success rates of these restorations can reduce the burden of dental diseases and associated health care costs, enhancing the quality of life for affected children. On the premise of this background, this systematic review and meta-analysis aims to provide a comprehensive qualitative assessment of crown-retention rates and the impact of technical and biological factors on crown success and acceptability for primary teeth.

Methods

Overview

The protocol for this systematic review and meta-analysis [14] was registered with PROSPERO (CRD42023442266) and conducted in accordance with the MOOSE (Meta-Analysis of Observational Studies in Epidemiology) and PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) guidelines [15,16].

Eligibility Criteria

RCTs and clinical studies (prospective and retrospective) evaluating crowns fabricated using stainless steel, zirconia, composite resin, compomer, and resin-modified glass ionomer cement (RMGIC) were included. The predefined inclusion criteria for this systematic review encompassed studies with an English abstract that evaluated crown restorations in pediatric patients aged 1-10 years, reporting crown-retention data, reasons for retentive loss, and technical and biological complications. Qualitative interviews, quasi-experimental studies, single-case studies, and series of single-case studies were excluded. Additionally, articles based on conference abstracts and dissertations were excluded.

Biological factors in this analysis included secondary caries and periodontal pathology (including the periodontal index, which assesses the severity of periodontal disease). Clinical parameters included the gingival index (evaluating gingival inflammation) and other periodontal health indicators. Technical factors included anatomic form, marginal adaptation, color match, surface texture, and wear on the opposing tooth. Success rates, if reported, were also included for the purposes of this systematic review and meta-analysis.

Data Sources and Search Strategy

Using a database-appropriate search strategy, electronic databases, including Cochrane, Embase, and PubMed (MEDLINE), were systematically searched to identify RCTs and clinical studies (prospective and retrospective) based on the eligibility criteria described above. The PICOS (Population, Intervention, Comparison, Outcomes, and Study design) format for this systematic review and meta-analysis was as follows:

- Population: children with primary tooth decay
- Intervention: crown restorations
- Comparators: materials such as stainless steel, zirconia, composite resin, compomer, and RMGIC
- Outcomes: crown retention, as well as technical and biological factors
- Study design: RCTs and clinical studies (prospective and retrospective)

Multimedia Appendix 1 presents the search syntax used for searching the electronic databases. The search strategy was considered adequate to reduce the risk of selection and detection bias.

Selection of Studies

The search results were imported into Zotero, and duplicates were removed to create a virtual library. Study selection was conducted in 2 stages. In the first stage, 2 independent reviewers (SL and KJ) screened the titles and abstracts of identified studies to determine their eligibility. In the second stage, the same 2 reviewers independently assessed the full-text articles of potentially relevant studies to confirm whether they met the eligibility criteria. Quality assessment of included RCTs and prospective and retrospective clinical studies was also

XSL•FO
RenderX

independently performed by SL and KJ. Any disagreements at any stage, including study selection, quality assessment, and data extraction, were resolved through discussion. If consensus could not be reached, a third expert (DG) was consulted, with DG making the final decision.

Data Extraction

Data extraction was independently conducted by 2 reviewers (SL and KJ) using a standardized data extraction form. Extracted data included author details, year of publication, number of children, age, number of teeth, tooth type, crown material, dropout, follow-up duration, crown retention, and retention complications including biological and technical complications. Discrepancies in data extraction were resolved through discussion between the 2 reviewers. If consensus could not be reached, a third expert (DG) was consulted to make the final decision.

Risk-of-Bias Assessments

The Risk of Bias 2 tool developed by the Cochrane Collaboration was used to assess the quality of included RCTs [17]. Two independent reviewers evaluated various bias domains in RCTs, including randomization, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other sources of bias. After assessing each included RCT for risk-of-bias domains specified by the Risk of Bias 2 tool, an overall risk-of-bias judgment (low, some concerns, or high) was assigned to each trial [17]. A checklist proposed by Moga et al [18] was adapted for assessing the risk of bias in the included prospective and retrospective clinical studies. Only studies with a moderate or low risk of bias were included in this analysis. Furthermore, a funnel plot was generated to assess for any publication bias.

Meta-Analytic Approach

For quantitative analysis (meta-analysis), retention was defined as the number of crowns that were in situ, regardless of technical and biological complications. Total exposure time refers to the cumulative duration that all crowns remain in the mouth across all included studies. It is calculated by summing the duration each crown is in place for each study and then aggregating these times across all studies. Failure rates resulting from retentive complications were calculated by dividing the number of failures by the total exposure time. Exposure time for each included study was calculated by summing the exposure time for all restorations. A Poisson regression model was used to analyze the calculated rates. Proportions of crowns retained at 3 years and 5 years were estimated with an assumption of constant event rates. The Pearson goodness-of-fit test was performed to assess heterogeneity across studies and to determine whether a fixed-effects or random-effects model should be used. Since no statistically significant heterogeneity was detected in any of the individual groups and rates (P>.05), a fixed-effects Poisson regression model was used to estimate the parameters. A P value <.05 was considered significant. All analyses were performed using R statistical software (version 4.1.2; R Core Team).

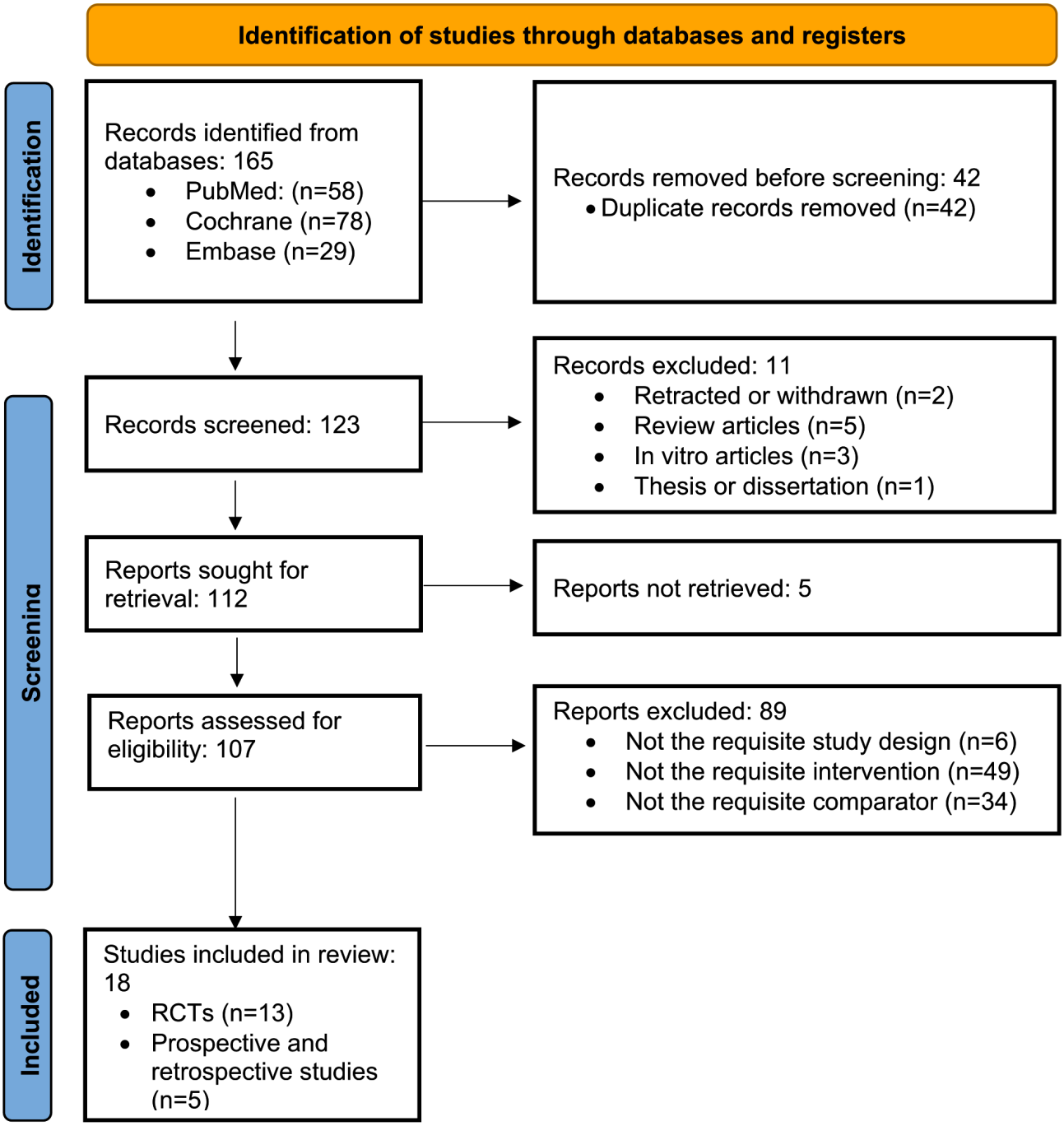
Results

Selection of Studies and Risk-of-Bias Assessments

A total of 13 RCTs and 5 clinical studies (prospective and retrospective) were included in this systematic review [19-31]. Figure 1 presents the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram illustrating the study selection process. Tables 1 and 2 provide the risk-of-bias assessments for the included RCTs and clinical studies, respectively.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for selection of studies included in the systematic review and meta-analysis. RCT: randomized controlled trial.



XSL•FO
RenderX

Table . Risk-of-bias assessments of the included randomized controlled trials using the Cochrane Risk of Bias tool.

Authors	Selection bias		Performance bias	Detection bias	Attrition bias	Reporting bias	Other bias
	Random sequence allotment	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Funding or conflicts of interest
Zulekha et al [19]	+++	+[a]	++[b]	+++[c]	+++	+++	+++
Vaghela et al [20]	+++	+	+	+	+++	+++	+++
Talekar et al [21]	+++	++	++	+++	+++	+++	+++
Hanafi et al [22]	+++	+++	+++	+++	+++	+++	+++
Güçlü et al [23]	+++	+	+	+	+++	+++	+++
Nischal et al [24]	+++	+	+	+	+++	+++	+++
Mathew et al [25]	+++	+	+	+	+++	+++	+++
Gill et al [26]	+++	++	++	+++	+++	+++	+++
Taran and Kaya [27]	+++	+	+	+	+++	+++	+++
Donly et al [28]	+++	++	++	+++	+++	+++	+++
Bektas Donmez et al [29]	+++	++	++	+++	+++	+++	+++
Sengul and Gurbuz [30]	+++	+	+	+	+++	+++	+++
Walia et al [31]	+++	+	+	+	+++	+++	+++

[a]+: high risk of bias.

[b]++: moderate risk of bias.

[c]+++: low risk of bias.

Table . Risk-of-bias assessments of the included prospective and retrospective clinical studies.

Authors	Domain 1: study design	Domain 2: population	Domain 3: intervention	Domain 4: outcomes	Domain 5: statistical analysis	Domain 6: results or conclusion	Domain 7: competing interests	Overall
Prabhu et al [32]	++[a]	++	++	++	++	++	++	++
Alhissan and Pani [33]	++	++	++	++	++	++	++	++
Holsinger et al [34]	++	++	++	++	++	++	–[b]	+[c]
Bücher et al [35]	++	++	++	++	++	++	++	++
Daou et al [36]	++	++	++	++	++	++	++	++

[a]++: low risk of bias.

[b]–: serious risk of bias.

[c]+: moderate risk of bias.

Qualitative Analysis of the Included Studies

Table 3 summarizes the 13 included RCTs, presenting data on crown retention and retentive complications [19-31]. These RCTs were conducted between 2014 and 2022 and recruited 454 children who received a total of 1172 crowns [19-31]. The follow-up duration ranged from 6 to 36 (median 12, IQR 9) months, with a cumulative dropout rate of 10.6% (124/1172

XSL•FO
RenderX

crowns). Table 4 outlines the 5 included clinical studies, reporting data on crown retention and retentive complications [32-36]. These clinical studies were conducted between 2008 and 2020 and included data from 810 children, with a follow-up duration ranging from 12 to 24 (median 20.8, IQR 5) months [32-36].

XSL•FO
RenderX

Table . Summary of included randomized controlled trials along with their crown-retention data.

Authors	Children, n	Age (years)	Tooth type	Number of teeth; crown material	Dropout, n (%)[a]	Follow-up duration (months)	Crown retention, n/N (%)	Retention complications
Zulekha et al [19]	25	3 to 5	Incisors	• 25; one shade composite resin • 25; composite resin	• 0 (0) • 0 (0)	• 12 • 12	• 25/25 (100) • 25/25 (100)	• None • None
Vaghela et al [20]	31	3 to 6	Incisors	• 55; strip crowns • 47; zirconia	• 4 (7) • 2 (4)	• 9 • 9	• 44/55 (80.4) • 46/47 (97.8)	• Chipping (4); material loss (6) • Complete loss of crown (1)
Talekar et al [21]	30	4 to 9	Molars	• 33; glass-reinforced resin • 33; zirconia	• 1 (3) • 1 (3)	• 18 • 18	• 29/33 (87.8) • 31/33 (93.9)	• Chipped (13); complete loss of crown (1) • None
Hanafi et al [22]	44	5 to 9	Molars and incisors	• 31; CCZC[b] • 32; NZC[c]	• 1 (3) • 1 (3)	• 6 • 6	• 31/31 (100) • 31/32 (96.9)	• None • Crown fracture (1)
Güçlü et al [23]	26	1 to 10	Molars and incisors	• 11; strip crowns • 22; zirconia • 13; stainless steel	• 4 (15) • 4 (15) • 4 (15)	• 6 • 6 • 6	• 11/11 (100) • 19/22 (86.4) • 13/13 (100)	• None • Decementation (3) • None
Nischal et al [24]	45	Not known	Incisors	• 15; strip crowns • 15; zirconia • 15; luxa	• Not known • Not known • Not known	• 9 • 9 • 9	• 12/15 (80) • 15/15 (100) • 12/15(80)	• Loss of bulk (2) • None • Loss of bulk (2)
Mathew et al [25]	30	6 to 8	Molars	• 30; zirconia • 30; stainless steel	• Not known • Not known	• 36 • 36	• 30/30 (100) • 30/30 (100)	• None • None
Gill et al [26]	49	2 to 4	Incisors	• 70; strip crowns • 70; zirconia • 80; stainless steel	• 22 (31) • 30 (43) • 33 (41)	• 12 • 12 • 12	• 55/70 (79) • 67/70 (95) • 80/80 (100)	• Partial material loss (6); complete material loss (3) • Partial material loss (1) • None

XSL•FO
RenderX

Authors	Children, n	Age (years)	Tooth type	Number of teeth; crown material	Dropout, n (%)[a]	Follow-up duration (months)	Crown retention, n/N (%)	Retention complications
Taran and Kaya [27]	13	6 to 9	Molars	• 26; zirconia • 26; stainless steel	• Not known • Not known	• 12 • 12	• 25/26 (98) • 24/26 (92.3)	• Decementation (2) • None
Donly et al [28]	50	3 to 7	Molars	• 50; zirconia • 50; stainless steel	• Not known • Not known	• 24 • 24	• 50/50 (100) • 50/50 (100)	• None • None
Bektas Donmez et al [29]	31	4 to 7	Molars	• 31; RMGIC[d] • 31; compomer • 31; composite resin	• 2 (6) • 1 (3) • 4 (13)	• 18 • 18 • 18	• 28/31 (90.3) • 31/31 (100) • 25/31 (80.6)	• Poor anatomic form (3) • None • Poor anatomic form (3)
Sengul and Gurbuz [30]	41	5 to 7	Molars	• 40; hybrid composite resin • 32; RMGIC • 36; compomer • 38; giomer	• 0 (0) • 0 (0) • 0 (0) • 0 (0)	• 24 • 24 • 24 • 24	• 37/40 (92.5) • 29/32 (90.3) • 28/36 (77.8) • 34/38 (89.5)	• Crown fracture (3) • Crown fracture (3) • Crown fracture (8) • Crown fracture (4)
Walia et al [31]	39	3 to 5	Incisors	• 43; strip crowns • 43; zirconia • 43; stainless steel	• 7 (16) • 5 (12) • 6 (14)	• 6 • 6 • 6	• 34/43 (78) • 43/43 (100) • 41/43 (95)	• Partial material loss (2); complete loss of crown (7) • None • Partial material loss (2)

[a]Percentages use the number of teeth as the denominator.

[b]CCZC: conventional chairside zirconia crown.

[c]NZC: novel zirconia crown.

[d]RMGIC: resin-modified glass ionomer cement.

XSL•FO
RenderX

Table . Summary of included prospective and retrospective studies along with their crown-retention data.

Authors	Children, n	Age (years)	Tooth type	Number of teeth; crown material	Follow-up duration (months)	Crown retention, n/N (%)	Complications
Prabhu et al [32]	60	6 to 10	Molars	• 30; stainless steel • 30; zirconia	• 24 • 24	• 30/30 (100) • 30/30 (100)	• None • None
Alhissan and Pani [33]	20	3 to 5	Incisors	• 70; zirconia	• 24	• 56/70 (80)	• Debonding (18.57%); failure without debonding (1.43%)
Holsinger et al [34]	18	Not reported	Incisors	• 57; zirconia	• 20.8*	• 55/57 (96)	• Complete loss of crown (4%)
Bücher et al [35]	667	Not known	Molars or incisors	• 2388; composite fillings	• 19	• 1977/2388 (82.8)	• Technical failures (8.3%)
Daou et al [36]	45	6 to 8	Molars	• 39; PMRC[a] • 37; RMGIC[b] • 35; HVGIC[c] • 38; amalgam	• 12 • 12 • 12 • 12	• 39/39 (100) • 37/37 (100) • 35/35 (100) • 38/38 (100)	• None • None • None • None

[a]PMRC: polyacid-modified resin composite.

[b]RMGIC: resin-modified glass ionomer cement.

[c]HVGIC: high-viscosity glass ionomer cement.

Figure 2 provides a descriptive analysis of retention rates for crowns fabricated with different materials, as reported in the included RCTs [19-31]. Retention rates varied from 77.8% to 100%, from 80.6% to 100%, and from 92.3% to 100% for compomer, composite resin, and stainless steel crowns, respectively. Additionally, retention rates for strip crowns and zirconia crowns ranged from 78% to 100% and from 86.4% to 100%, respectively. Table 5 illustrates various retentive complications reported in the RCTs for different materials. Regarding clinical studies, debonding and loss of crown were reported for zirconia crowns, while crown fractures were reported for composite resin crowns (Table 4).

XSL•FO
RenderX

Figure 2. Retention rates for crowns fabricated with different materials, as reported in the included randomized controlled trials.

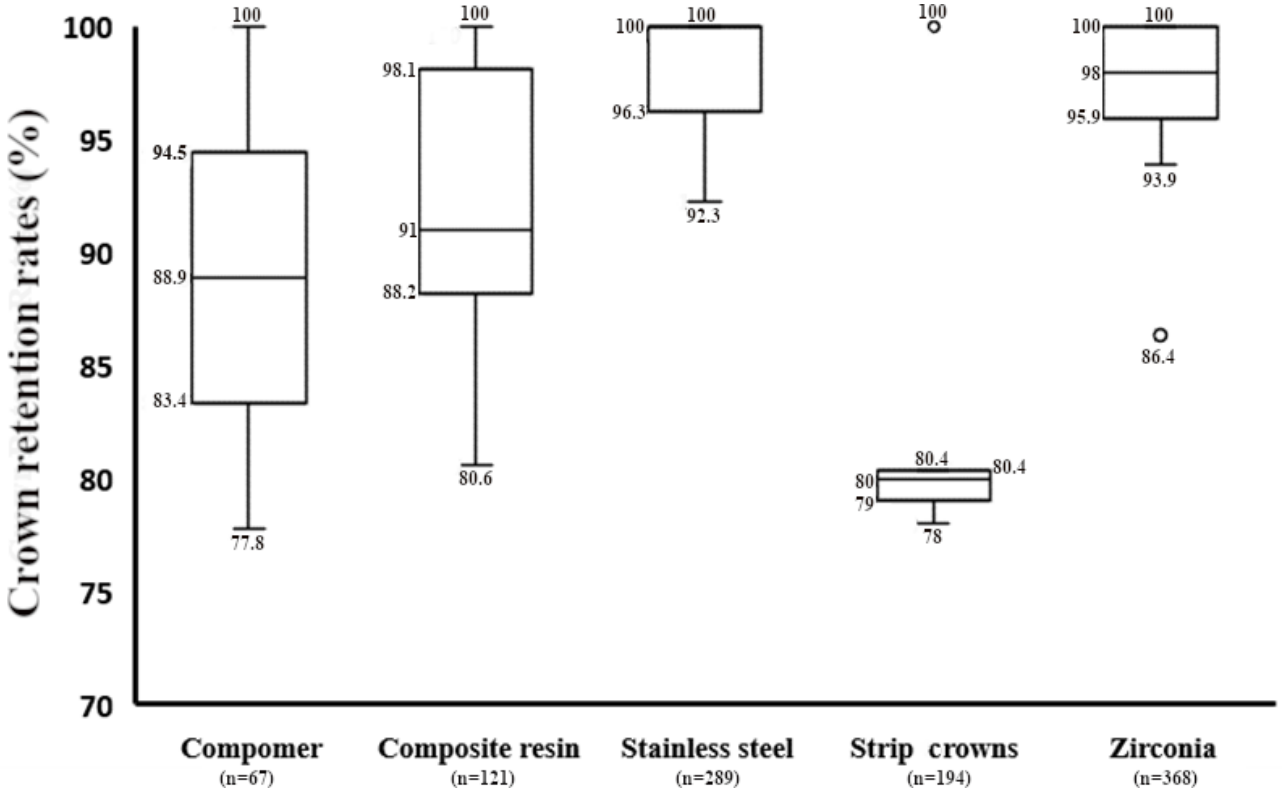


Table . Retentive complications reported in the randomized controlled trials for different materials.

Crown material	Decementation, n	Chipping[a], n	Crown fracture[b], n	Poor anatomic form, n	Complete crown loss, n	Partial material loss[c], n
Compomer	0	0	8	0	0	0
Composite resin	0	12	3	3	1	2
Stainless steel	0	0	0	0	0	0
Strip crowns	0	3	0	0	7	19
Zirconia	5	0	0	0	1	1

[a]Chipping: loss of small fragments or pieces from the surface of the crown.

[b]Crown fracture: severe damage where the crown splits or breaks into two or more parts.

[c]Partial material loss: a more substantial loss of crown material compared to chipping but does not constitute a complete fracture.

Biological Complications Reported in the Included Studies

Table 6 depicts the distribution of biological complications reported in the included RCTs [18-30]. In the prospective and retrospective clinical studies, 8.3% (6/67) of children receiving compomer crowns and 8.8% (11/121) receiving composite resin crowns reported secondary caries [31-35]. Additionally, 3.3% (10/289) and 21.8% (80/368) of children with stainless steel crowns and zirconia crowns, respectively, in these studies exhibited gingival inflammation [32-36].

XSL•FO
RenderX

Table . Biological and clinical complications reported in the randomized controlled trials for different materials.

Biological and clinical complications	Compomer (n=67)	Composite resin (n=121)	Stainless steel (n=289)	Strip crowns (n=194)	Zirconia (n=368)
Secondary caries, n (%)	4 (5.97)	3 (2.47)	13 (4.54)	12 (6.18)	0 (0)
Gingival inflammation, n (%)	0 (0)	29 (24.24)	23 (7.85)	14 (7.21)	13 (3.5)
Plaque index[a], n (%)	0 (0)	0 (0)	0 (0)	0 (0)	25 (6.76)
Bleeding on probing, n (%)	0 (0)	0 (0)	0 (0)	0 (0)	43 (11.77)

[a]Plaque Index: a clinical measure used to assess the amount of dental plaque on teeth.

Technical Complications Reported in the Included Studies

Table 7 outlines the distribution of technical complications reported in the included RCTs [18-30]. In the prospective and retrospective clinical studies, 29% (19/67) of compomer crowns exhibited marginal adaptation complications, while stainless steel crowns did not demonstrate marginal adaptation issues [32-36]. No marginal discoloration or plaque retention complications were observed for compomer crowns. Among the 399 zirconia crowns evaluated, 4.8% (19/399) demonstrated shade mismatches, and 1% (4/399) exhibited marginal integrity complications [32-36].

Table . Percentage of technical complications reported in the included randomized controlled trials [19-31].

Crown material	Surface roughness	Occlusal wear	Marginal adaptation	Marginal discoloration	Marginal integrity	Staining	Plaque retention	Shade mismatch	Opposing tooth wear
Compomer (n=67), n (%)	0 (0)	0 (0)	19 (29)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Composite resin[a] (n=154), n (%)	2 (1.3)	17 (11)	8 (5.2)	2 (1.3)	0 (0)	24 (15.6)	20 (13)	45 (29.2)	0 (0)
Stainless steel (n=242), n (%)	0 (0)	0 (0)	0 (0)	0 (0)	1 (0.4)	7 (2.9)	0 (0)	0 (0)	33 (13.6)
Strip crowns (n=194), n (%)	0 (0)	0 (0)	14 (7.2)	0 (0)	7 (3.6)	6 (3.1)	0 (0)	41 (21.11)	31 (16)
Zirconia (n=399), n (%)	1 (0.3)	0 (0)	2 (0.5)	0 (0)	4 (1)	7 (1.8)	2 (0.5)	19 (4.8)	34 (8.5)

[a]Technical complications were reported more for composite resin in the included studies.

Quantitative Analysis (Meta-Analysis) of the Included Studies

Table 8 presents the estimated 3-year and 5-year retention rates of crowns fabricated using various materials, as reported in the included RCTs [19-31].

The Pearson goodness-of-fit test was performed to assess heterogeneity across studies and to determine whether a fixed-effects or random-effects model should be used. Since no statistically significant heterogeneity was detected in any of the individual groups and rates (P>.05), a fixed-effects Poisson regression model was used to estimate the parameters.

Meta-analysis of pooled 5-year retention rates for different crown materials revealed the following retention rates: 88.90% for compomer crowns (Figure 3), 92.18% for composite resin crowns (Figure 4), 90.30% for RMGIC crowns (Figure 5), and 97.88% for stainless steel crowns (Figure 6). Additionally, strip crowns exhibited a retention rate of 83.48% (Figure 7), while zirconia crowns had a retention rate of 97.09% (Figure 8). Figure 9 presents the retention rates, irrespective of the material, and found a retention rate of 92.20% (95% CI 92.14% - 92.26%).

XSL•FO
RenderX

Table . Estimated 3-year and 5-year crown-retention rates in primary tooth from data reported in the included randomized controlled trials [19-31].

Authors and crown material	Exposure time (days), n	Estimated failure rate per crown year	Estimated 3-year retention rate (%)	Estimated 5-year retention rate (%)
Zulekha et al [19]				
One-shade composite resin	300	0	100	100
Composite resin	300	0	100	100
Walia et al [31]				
Strip crowns	258	0.42	28.5	12.3
Zirconia	258	0	100	100
Stainless steel	258	0.09	75.7	62.8
Vaghela et al [20]				
Strip crowns	495	0.24	48.3	29.8
Zirconia	423	0.03	91.8	86.8
Taran and Kaya [27]				
Stainless steel	312	0	100	100
Zirconia	312	0.08	79.3	68.1
Talekar et al [21]				
Glass-reinforced resin	594	0.28	42.8	24.3
Zirconia	594	0	100	100
Sengul and Gurbuz [30]				
Hybrid composite resin	960	0.04	89.4	82.9
RMGIC[a]	768	0.05	86.9	79.1
Compomer	864	0.11	71.7	57.4
Giomer	912	0.05	85.3	76.9
Nischal et al [24]				
Strip crowns	135	0.18	58.7	41.1
Zirconia	135	0	100	100
Luxa	135	0.18	58.7	41.1
Mathew et al [25]				
Zirconia	1080	0	100	100
Stainless steel	1080	0	100	100
Hanafi et al [22]				
Zirconia	186	0	100	100
Zirconia	192	0.06	82.9	73.2
Güçlü et al [23]				
Strip crowns	66	0	100	100
Zirconia	132	0.27	44.1	25.6
Stainless steel	78	0	100	100
Gill et al [26]				
Strip crowns	840	0.13	68	52.6
Zirconia	840	0.01	95.8	93.1
Stainless steel	960	0	100	100
Bektas Donmez et al [29]				
RMGIC	558	0.06	82.4	72.4

XSL•FO
RenderX

Authors and crown material	Exposure time (days), n	Estimated failure rate per crown year	Estimated 3-year retention rate (%)	Estimated 5-year retention rate (%)
Compomer	558	0	100	100
Composite resin	558	0.06	82.4	72.4
Donly et al [28]				
Zirconia	1200	0	100	100
Stainless steel	1200	0	100	100

[a]RMGIC: resin-modified glass ionomer cement.

Figure 3. Forest plot illustrating the retention rates of compomer crowns [29,30].

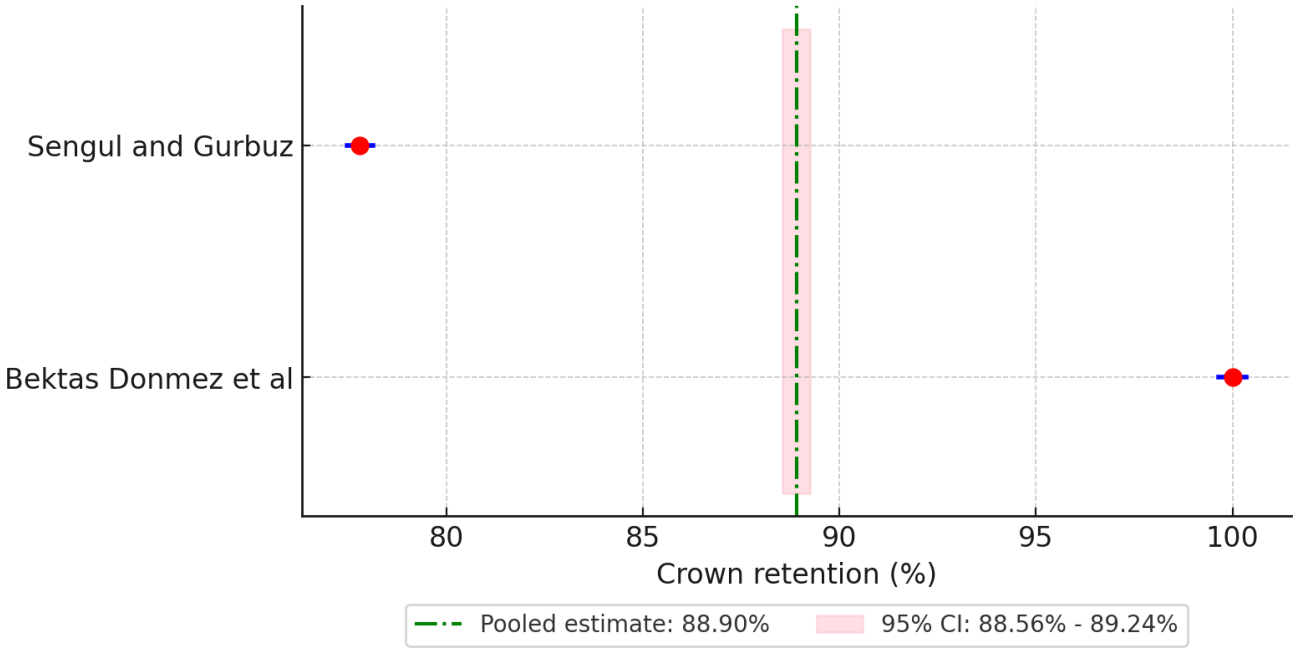
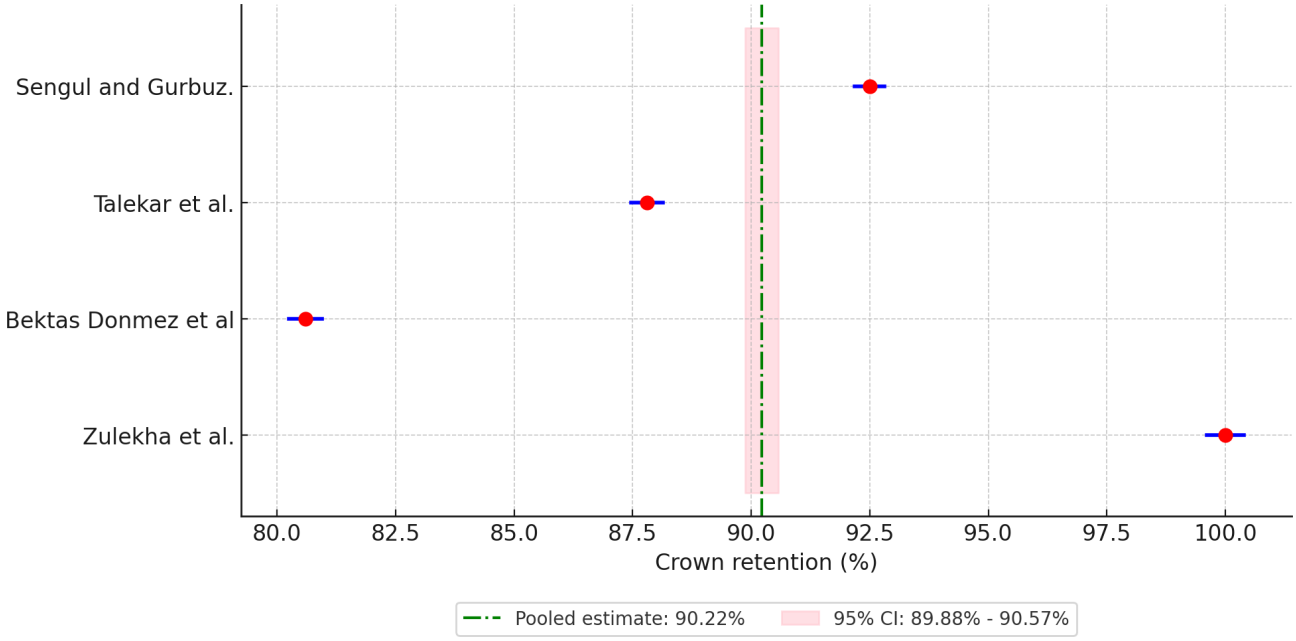


Figure 4. Forest plot illustrating the retention rates of composite resin crowns [19,21,29,30].



XSL•FO
RenderX

Figure 5. Forest plot illustrating the retention rates of resin-modified glass ionomer cement (RMGIC) crowns [29,30].

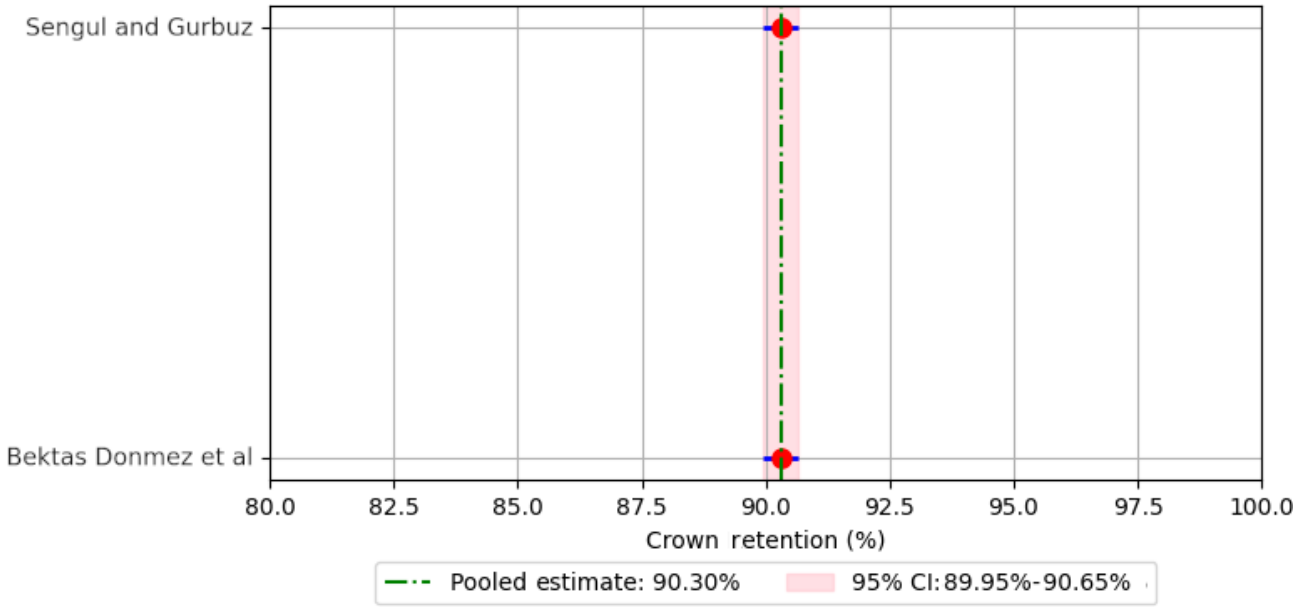
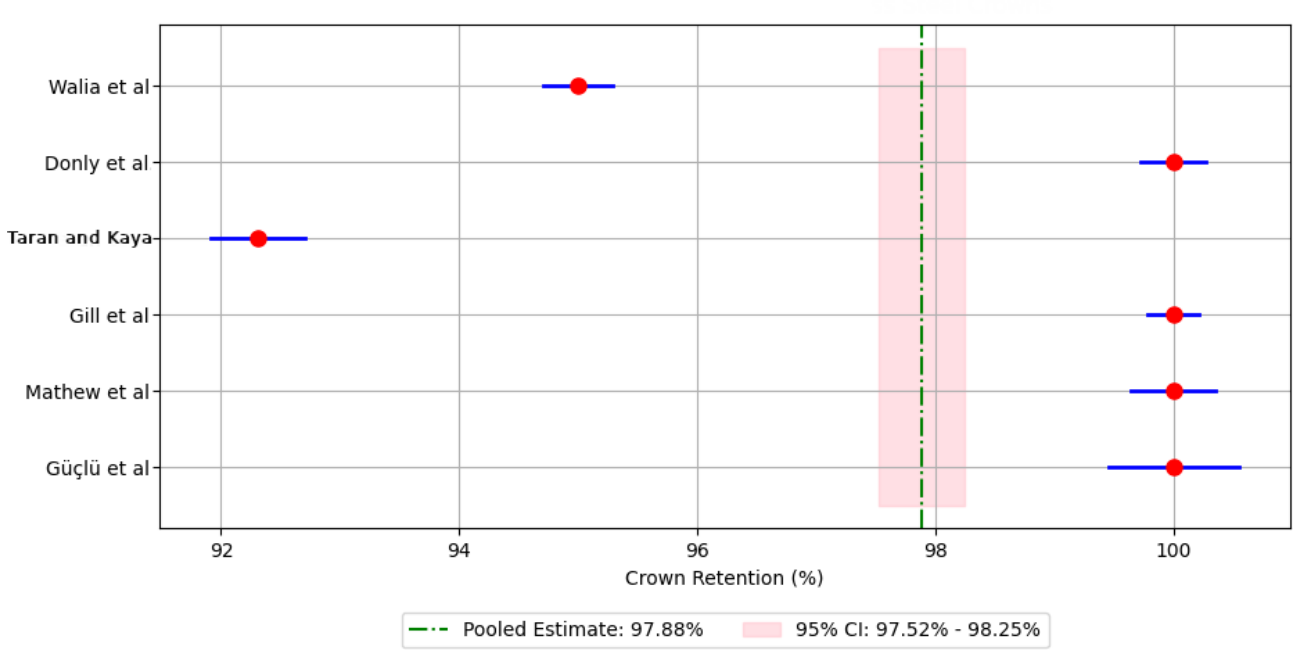


Figure 6. Forest plot illustrating the retention rates of stainless steel crowns [23,25-28,31].



XSL•FO
RenderX

Figure 7. Forest plot illustrating the retention rates for strip crowns [20,23,24,26,31].

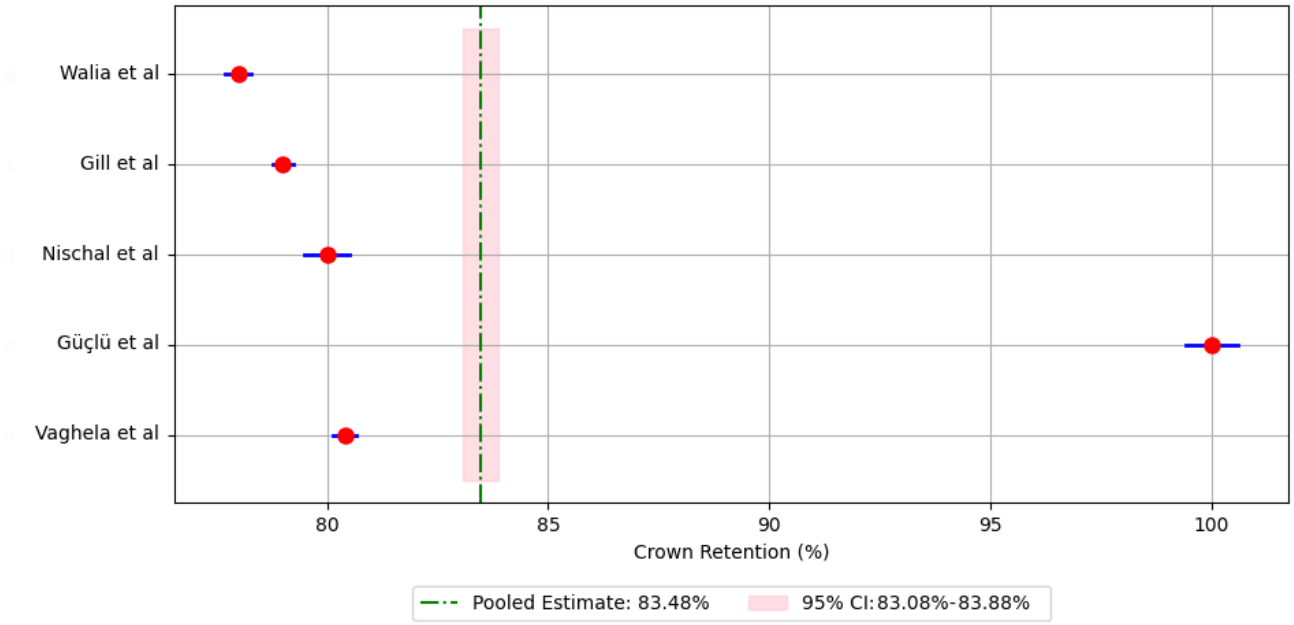
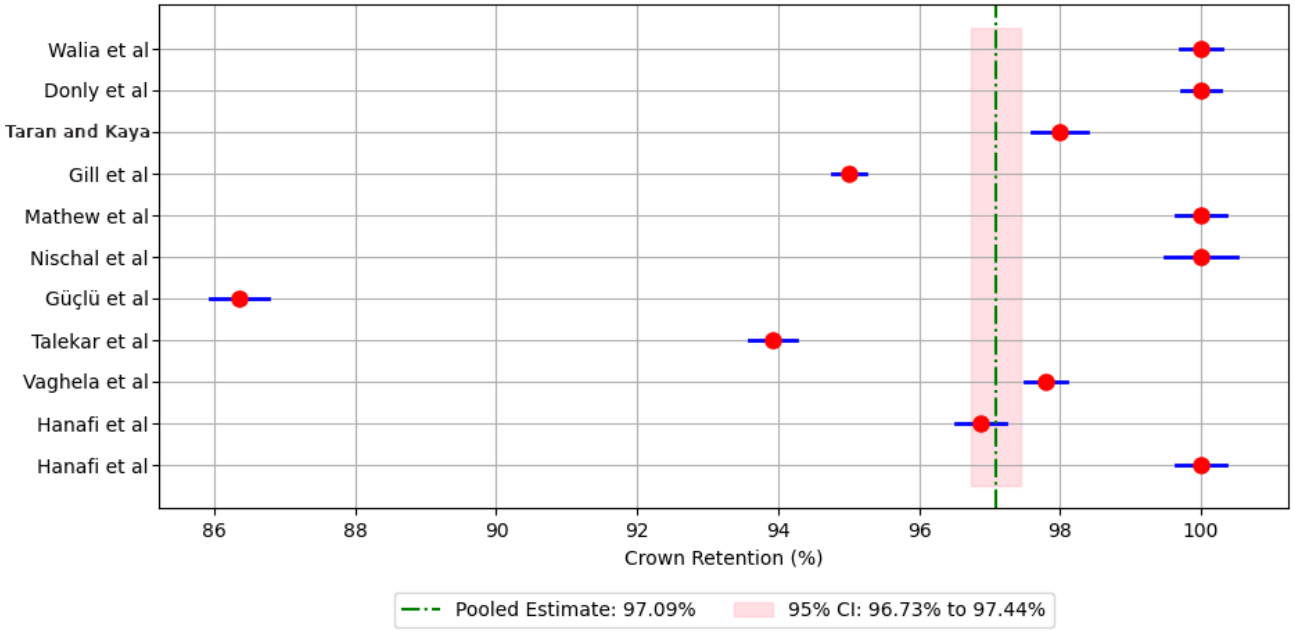
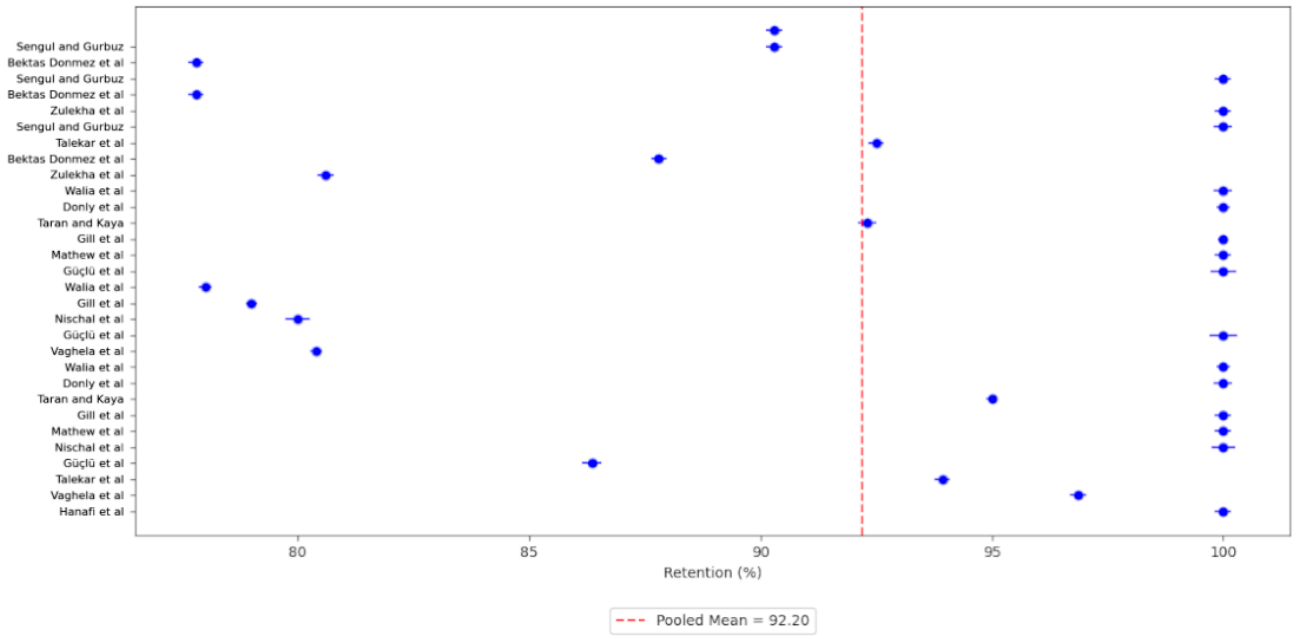


Figure 8. Forest plot illustrating the retention rates for zirconia crowns [20-28,31].



XSL•FO
RenderX

Figure 9. Forest plot illustrating pooled 5-year retention rates across all studies without stratification by crown material [19-31].



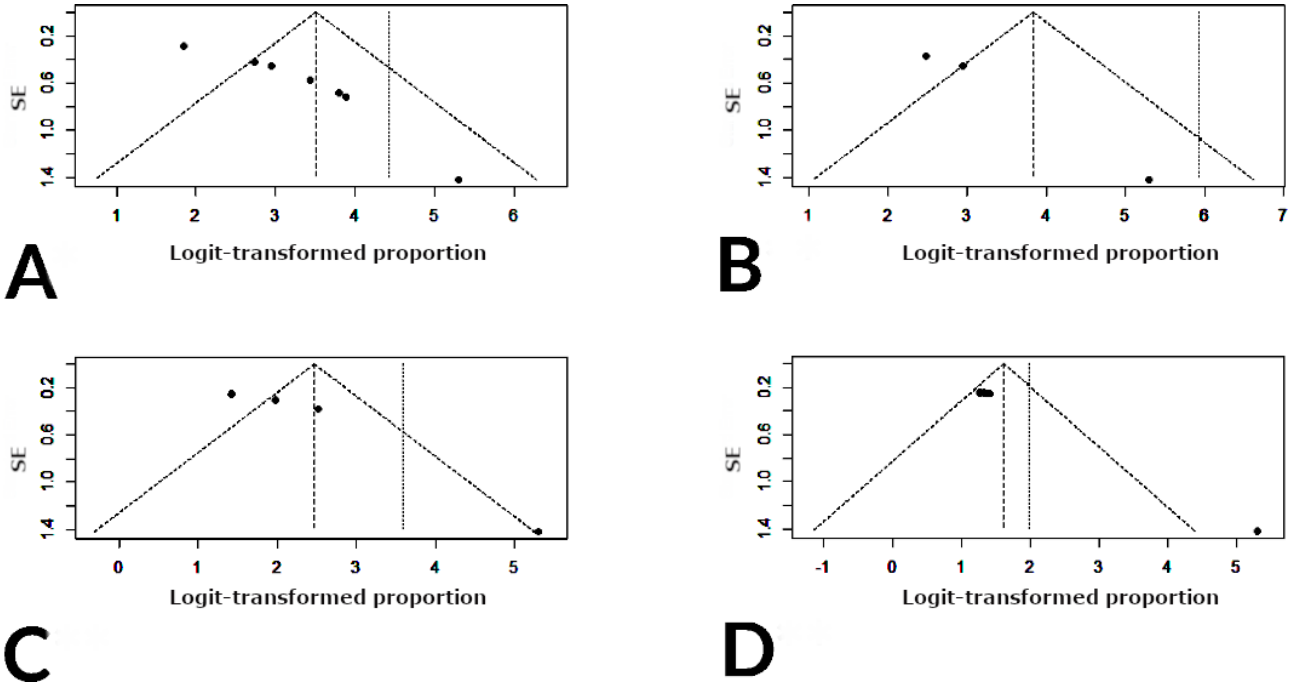
Discussion

This systematic review examined data from RCTs and prospective and retrospective clinical studies to evaluate the performance of dental crowns for primary teeth. Key findings include significant biases in RCTs affecting the assessment of crown efficacy, variability in age distribution impacting retention rates, and varying complications across different crown materials. Despite these challenges, stainless steel crowns consistently demonstrated superior retention rates, while complications such as secondary caries and gingival inflammation varied among materials.

In our analysis, stainless steel crowns exhibited the highest retention rates, corroborating findings from previous studies [8,10], which also highlighted the superior performance of stainless steel crowns in terms of retention and durability. The study by Chisini et al [8] indicated an annual failure rate between 0% and 29.9%, with stainless steel crowns showing the highest success rate of 96.1%. In this study, the estimated 3-year and 5-year retention rates ranged from 75.7% to 100% and from 62.8% to 100%, respectively. Zirconia and composite resin crowns also demonstrated high retention rates, although they were associated with different types of complications. These results align with another study by Alzanbaqi et al [9], which noted improved gingival and periodontal health, excellent retention, and high fracture resistance with zirconia crowns for primary teeth. Regarding complications, strip and composite resin crowns showed higher instances of chipping and partial material loss, likely due to the inherent brittleness of these materials and their susceptibility to fracture under masticatory forces [37-39]. Crowns fabricated with composite resins and newer materials like compomer exhibited a higher incidence of fractures, possibly due to inadequate bonding strength and less favorable mechanical properties compared with stainless steel and zirconia [39,40]. Composite resin crowns also exhibited higher rates of gingival inflammation, likely due to plaque accumulation and potential marginal leakage. Conversely, zirconia crowns showed minimal gingival inflammation, attributed to their smooth surface and biocompatibility [40-42]. Secondary caries was a significant issue across all materials but was particularly prevalent in crowns with poor marginal adaptation, emphasizing the need for precise placement and proper maintenance to mitigate caries risk [42,43]. While the findings demonstrate the efficacy of stainless steel crowns, material choice should consider the specific needs and conditions of pediatric patients. Clinicians must weigh the benefits of high retention rates against the potential for biological complications. The included RCTs displayed a notable prevalence of allocation concealment, performance bias, and detection bias (Table 1) [19-31]. These biases, inherent to the nature of interventions that often preclude optimal double blinding, necessitate careful consideration in assessing the risk of bias for studies involving crowns for primary teeth. These findings align with a previous systematic review by Chisini et al [8], which also identified a high risk of bias. The assessment of publication bias in our meta-analysis revealed noteworthy findings (Figure 10). The Egger test for zirconia crowns suggested potential publication bias. In contrast, for stainless steel, composite resin, and strip crowns, the limited number of studies precluded robust evaluation of small study effects. Thus, caution is warranted in interpreting these outcomes due to the small sample size. The significant results indicate asymmetry in the funnel plot, potentially influencing effect size estimates.

XSL•FO
RenderX

Figure 10. Funnel plots for assessments of publication bias in the included studies for (A) zirconia, (B) stainless steel, (C) composite resin, and (D) strip crowns. For part A, the Egger test was performed (t_9=9.07, P<.001). For parts B-D, the number of studies was below the recommended threshold for robustly assessing publication bias, and the potential for small study effects should be considered with prudence.



The age distribution of the study population is often not reported, posing challenges in determining optimal retention rates for a smooth transition from primary to permanent teeth, especially in younger age groups. The age of children in the RCTs and prospective and retrospective clinical studies ranged from 1 to 10 years and from 3 to 10 years, respectively [19-31]. The reported crown-retention rates for different age groups suggest potential nuances in the transition from primary to permanent teeth, underscoring the need for future studies to provide detailed age-distribution data. Another major limitation of this review is the varying sample sizes across studies, which could influence the generalizability of the results. However, the application of Poisson regression allowed for the adjustment of these differences, providing a more accurate comparison of retention rates [44]. Despite these limitations, this review provides a comprehensive analysis of crown retention and associated complications in primary teeth, offering valuable insights for evidence-based clinical decision-making.

Future studies should aim to include larger sample sizes and longer follow-up periods to validate the findings of this review. Additionally, research should focus on the development and evaluation of new materials and techniques that could enhance crown retention and reduce complications. Investigating advanced fabrication methods such as 3D printing and computer-aided design and computer-aided manufacturing technology holds promise for improving the fit and longevity of dental crowns.

In conclusion, while stainless steel crowns remain a gold standard for primary teeth restorations due to their high retention rates and durability, the choice of material should be tailored to individual patient needs, considering both mechanical properties and potential biological complications. Ongoing research and clinical advancements will continue to refine these treatment modalities, ultimately improving outcomes for pediatric dental patients.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search syntax used for searching databases.
[DOCX File, 13 KB - i-jmr_v14i1e57958_app1.docx]

Checklist 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist 2020.
[PDF File, 62 KB - i-jmr_v14i1e57958_app2.pdf]

References

XSL•FO
RenderX

1. Kassebaum NJ, Bernabé E, Dahiya M, Bhandari B, Murray CJL, Marcenes W. Global burden of untreated caries: a systematic review and metaregression. J Dent Res 2015 May;94(5):650-658. [doi: 10.1177/0022034515573272] [Medline: 25740856]
2. Manfredini D, Restrepo C, Diaz-Serrano K, Winocur E, Lobbezoo F. Prevalence of sleep bruxism in children: a systematic review of the literature. J Oral Rehabil 2013 Aug;40(8):631-642. [doi: 10.1111/joor.12069] [Medline: 23700983]
3. Wetselaar P, Vermaire EJH, Lobbezoo F, Schuller AA. The prevalence of awake bruxism and sleep bruxism in the Dutch adolescent population. J Oral Rehabil 2021 Feb;48(2):143-149. [doi: 10.1111/joor.13117] [Medline: 33070349]
4. Chengappa MM D, Kannan A, Sharma D. Anterior and posterior crowns in primary dentition: a contemporary review. Int J Oral Care Res 2020;8(4):83-87. [doi: 10.4103/INJO.INJO_35_20]
5. Dhar V, Hsu KL, Coll JA, et al. Evidence-based update of pediatric dental restorative procedures: dental materials. J Clin Pediatr Dent 2015;39(4):303-310. [doi: 10.17796/1053-4628-39.4.303] [Medline: 26161599]
6. Amlani DV, Brizuela M. Stainless steel crowns in primary dentition. In: StatPearls: StatPearls Publishing; 2024. URL: https://www.ncbi.nlm.nih.gov/books/NBK574547/ [accessed 2025-04-17]
7. Vollmer Dahlke D, Fair K, Hong YA, Beaudoin CE, Pulczinski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. JMIR Mhealth Uhealth 2015 Mar 27;3(1):e31. [doi: 10.2196/mhealth.3861] [Medline: 25830810]
8. Chisini LA, Collares K, Cademartori MG, et al. Restorations in primary teeth: a systematic review on survival and reasons for failures. Int J Paediatr Dent 2018 Mar;28(2):123-139. [doi: 10.1111/ipd.12346] [Medline: 29322626]
9. Alzanbaqi SD, Alogaiel RM, Alasmari MA, et al. Zirconia crowns for primary teeth: a systematic review and meta-analyses. Int J Environ Res Public Health 2022 Feb 28;19(5):2838. [doi: 10.3390/ijerph19052838] [Medline: 35270531]
10. Gurunathan D, Suguna S. Clinical and radiographic evaluation of stainless steel crowns and zirconia crowns in primary molars: a systematic review. Ann Med Health Sci Res 2021 Jul 2 [FREE Full text]
11. Chua DR, Tan BL, Nazzal H, Srinivasan N, Duggal MS, Tong HJ. Outcomes of preformed metal crowns placed with the conventional and Hall techniques: a systematic review and meta-analysis. Int J Paediatr Dent 2023 Mar;33(2):141-157. [doi: 10.1111/ipd.13029] [Medline: 36151937]
12. Elheeny AAH, Abdelmotelb MA. Oral health-related quality of life (OHRQOL) of preschool children's anterior teeth restored with zirconia crowns versus resin-bonded composite strip crowns: a 12-month prospective clinical trial. Clin Oral Investig 2022 May;26(5):3923-3938. [doi: 10.1007/s00784-021-04359-9] [Medline: 34989861]
13. Almajed OS. Shaping smiles: a narrative review of crown advancements in pediatric dentistry. Cureus 2024 Jan;16(1):e52997. [doi: 10.7759/cureus.52997] [Medline: 38406007]
14. Lampl S, Gurunathan D, Krithikadatta J, Mehta D, Moodley D. Reasons for crown failures in primary teeth: protocol for a systematic review and meta-analysis. JMIR Res Protoc 2023 Nov 1;12:e51505. [doi: 10.2196/51505] [Medline: 37910174]
15. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. JAMA 2000 Apr 19;283(15):2008-2012. [doi: 10.1001/jama.283.15.2008] [Medline: 10789670]
16. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
17. 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]
18. Moga C, Guo B, Schopflocher D, Harstall C. Development of a Quality Appraisal Tool for Case Series Studies Using a Modified Delphi Technique: Institute of Health Economics; 2012. URL: https://www.ihe.ca/advanced-search/development-of-a-quality-appraisal-tool-for-case-series-studies-using-a-modified-delphi-technique [accessed 2024-02-19]
19. Zulekha, Vinay C, Uloopi KS, RojaRamya KS, Penmatsa C, Ramesh MV. Clinical performance of one shade universal composite resin and nanohybrid composite resin as full coronal esthetic restorations in primary maxillary incisors: a randomized controlled trial. J Indian Soc Pedod Prev Dent 2022;40(2):159-164. [doi: 10.4103/jisppd.jisppd_151_22] [Medline: 35859408]
20. Vaghela LL, Patel MC, Bhatt RK, Patel CN, Joshi KR. Clinical performance and parental satisfaction with composite strip crown and prefabricated zirconia crown for primary anterior teeth: a randomized clinical trial. J Contemp Dent Pract 2021 Dec 1;22(12):1462-1470. [Medline: 35656688]
21. Talekar AL, Chaudhari GS, Waggoner WF, Chunawalla YK. An 18-month prospective randomized clinical trial comparing zirconia crowns with glass-reinforced fiber composite crowns in primary molar teeth. Pediatr Dent 2021 Sep 15;43(5):355-362. [Medline: 34654496]
22. Hanafi L, Altinawi M, Comisi JC. Evaluation and comparison two types of prefabricated zirconia crowns in mixed and primary dentition: a randomized clinical trial. Heliyon 2021 Feb;7(2):e06240. [doi: 10.1016/j.heliyon.2021.e06240] [Medline: 33665422]
23. Güçlü ZA, Çalışkan S, Efe Z, Doğan S. Can zirconia crowns be the first restorative choice after endodontic treatment of primary teeth? Int J Clin Pract 2021 Dec;75(12):e14888. [doi: 10.1111/ijcp.14888] [Medline: 34536960]
24. Nischal M, Gupta T, Mehra M, Sadana G. Clinical comparison of three tooth-colored full-coronal restorations in primary maxillary incisors. Int J Clin Pediatr Dent 2020;13(6):622-629. [doi: 10.5005/jp-journals-10005-1842] [Medline: 33976486]

XSL•FO
RenderX

25. Mathew MG, Roopa KB, Soni AJ, Khan MM, Kauser A. Evaluation of clinical success, parental and child satisfaction of stainless steel crowns and zirconia crowns in primary molars. J Family Med Prim Care 2020 Mar;9(3):1418-1423. [doi: 10.4103/jfmpc.jfmpc_1006_19] [Medline: 32509626]
26. Gill A, Garcia M, Won An S, Scott J, Seminario AL. Clinical comparison of three esthetic full-coverage restorations in primary maxillary incisors at 12 months. Pediatr Dent 2020 Sep 15;42(5):367-372. [Medline: 33087221]
27. Taran PK, Kaya MS. A comparison of periodontal health in primary molars restored with prefabricated stainless steel and zirconia crowns. Pediatr Dent 2018 Sep 15;40(5):334-339. [Medline: 30355428]
28. Donly KJ, Sasa I, Contreras CI, Mendez MJC. Prospective randomized clinical trial of primary molar crowns: 24-month results. Pediatr Dent 2018 Jul 15;40(4):253-258. [Medline: 30345963]
29. Bektas Donmez S, Uysal S, Dolgun A, Turgut MD. Clinical performance of aesthetic restorative materials in primary teeth according to the FDI criteria. Eur J Paediatr Dent 2016 Sep;17(3):202-212. [Medline: 27759409]
30. Sengul F, Gurbuz T. Clinical evaluation of restorative materials in primary teeth class II lesions. J Clin Pediatr Dent 2015;39(4):315-321. [doi: 10.17796/1053-4628-39.4.315] [Medline: 26161601]
31. Walia T, Salami AA, Bashiri R, Hamoodi OM, Rashid F. A randomised controlled trial of three aesthetic full-coronal restorations in primary maxillary teeth. Eur J Paediatr Dent 2014 Jun;15(2):113-118. [Medline: 25102458]
32. Prabhu D, Anantharaj A, Praveen P, Rani SP, Sudhir R. A clinical and radiographic comparative evaluation of custom-made zirconia crowns using CAD-CAM and stainless steel crowns in primary molars. J Indian Soc Pedod Prev Dent 2022;40(1):34-42. [doi: 10.4103/jisppd.jisppd_269_21] [Medline: 35439881]
33. Alhissan AS, Pani SC. Factors influencing the survival of preformed zirconia crowns in children treated under general anesthesia. Int J Dent 2021;2021:5515383. [doi: 10.1155/2021/5515383] [Medline: 33833801]
34. Holsinger DM, Wells MH, Scarbecz M, Donaldson M. Clinical evaluation and parental satisfaction with pediatric zirconia anterior crowns. Pediatr Dent 2016;38(3):192-197. [Medline: 27306242]
35. Bücher K, Metz I, Pitchika V, Hickel R, Kühnisch J. Survival characteristics of composite restorations in primary teeth. Clin Oral Investig 2015 Sep;19(7):1653-1662. [doi: 10.1007/s00784-014-1389-9] [Medline: 25547072]
36. Daou MH, Tavernier B, Meyer JM. Clinical evaluation of four different dental restorative materials: one-year results. Schweiz Monatsschr Zahnmed 2008;118(4):290-295. [Medline: 18491670]
37. Kayal S, Kang YH, Finkelman M, Swee G, Loo CY. Retention of zirconia crowns compared to stainless steel crowns: an ex-vivo study. Pediatr Dent 2023 Mar 15;45(2):142-146. [Medline: 37106542]
38. Vignesh KC, Kandaswamy E, Muthu MS. A comparative evaluation of fracture toughness of composite resin vs protemp 4 for use in strip crowns: an in vitro study. Int J Clin Pediatr Dent 2020;13(1):57-60. [doi: 10.5005/jp-journals-10005-1711] [Medline: 32581481]
39. Toparli M, Aksoy T. Fracture toughness determination of composite resin and dentin/composite resin adhesive interfaces by laboratory testing and finite element models. Dent Mater 1998 Jul;14(4):287-293. [doi: 10.1016/s0109-5641(98)00041-4] [Medline: 10379258]
40. Prabhakar AR, Raj S, Raju OS. Comparison of shear bond strength of composite, compomer and resin modified glass ionomer in primary and permanent teeth: an in vitro study. J Indian Soc Pedod Prev Dent 2003 Sep;21(3):86-94. [Medline: 14703213]
41. Pei SL, Chen MH. Comparison of periodontal health of primary teeth restored with zirconia and stainless steel crowns: a systemic review and meta-analysis. J Formos Med Assoc 2023 Feb;122(2):148-156. [doi: 10.1016/j.jfma.2022.08.015] [Medline: 36180321]
42. el-Kalla IH. Marginal adaptation of compomers in class I and V cavities in primary molars. Am J Dent 1999 Feb;12(1):37-43. [Medline: 10477997]
43. Askar H, Tu YK, Paris S, Yeh YC, Schwendicke F. Risk of caries adjacent to different restoration materials: systematic review of in situ studies. J Dent 2017 Jan;56:1-10. [doi: 10.1016/j.jdent.2016.09.011] [Medline: 27697582]
44. Kianifard F, Gallo PP. Poisson regression analysis in clinical research. J Biopharm Stat 1995 Mar;5(1):115-129. [doi: 10.1080/10543409508835101] [Medline: 7613557]

Abbreviations

MOOSE: Meta-Analysis of Observational Studies in Epidemiology
PICOS: Population, Intervention, Comparison, Outcomes, and Study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
RCT: randomized controlled trial
RMGIC: resin-modified glass ionomer cement

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 01.03.24; peer-reviewed by N Kaur, V Venkataiah; revised version received 24.02.25; accepted 24.02.25; published 01.05.25.

Please cite as:
Lampl S, Gurunathan D, Mehta D, Jogikalmat K
Reasons for Crown Failures in Primary Teeth: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e57958
URL: https://www.i-jmr.org/2025/1/e57958
doi:10.2196/57958

© Stephan Lampl, Deepa Gurunathan, Deepak Mehta, Krithikadatta Jogikalmat. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 1.5.2025. 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

The Effect of Combining mHealth and Health Professional–Led Intervention for Improving Health-Related Outcomes in Chronic Diseases: Systematic Review and Meta-Analysis

Masashi Kanai[1,2*], PhD; Takahiro Miki[1*], PhD; Toshiya Sakoda[1], MSc; Yuta Hagiwara[1], MSc

[1]PREVENT Inc, Aichi, Japan

[2]Institute of Transdisciplinary Science for Innovation, Kanawaza University, Kanazawa, Japan

[*]these authors contributed equally

Corresponding Author:
Takahiro Miki, PhD
PREVENT Inc
IKKO Shinsakae Building 9F
1 Aoi, Higashi-ku
Aichi, 461-0004
Japan
Phone: 81 527157955
Email: miki.takahiro@prevent.co.jp

Abstract

Background: Chronic diseases such as diabetes and cardiovascular disease are global health challenges, affecting millions of people worldwide. Traditional health care often falls short in chronic disease management. This has led to the exploration of innovative solutions, such as mobile health (mHealth) technologies. mHealth, which leverages mobile and wireless technologies, has the potential to transform health care delivery by providing continuous, accessible, and personalized care. However, the effectiveness of mHealth, particularly when integrated with traditional health care interventions delivered by professionals, warrants comprehensive investigation. Understanding the combined impact of mHealth and professional-led interventions is critical to maximizing the potential of mHealth to improve patient outcomes and adherence.

Objective: This study aims to investigate the effectiveness of combining mHealth and health professional–led intervention for improving health-related outcomes in chronic diseases

Methods: This systematic review and meta-analysis focused on randomized controlled trials. We searched Web of Science, CENTRAL, MEDLINE, and CINAHL through July 17, 2023. The study targeted patients aged 18 years and older, experiencing at least 1 chronic condition. The interventions were a combination of mHealth and the use of a health care professional. The comparison groups consisted of participants receiving either general care and follow-up or those using mHealth devices without any health care professional involvement. The outcomes measured in this review included hemoglobin A_{1c} (HbA_{1c}), quality of life (QoL), and physical activity.

Results: The study included 26 research papers, encompassing 7360 individuals. Meta-analysis was conducted for HbA_{1c}, QoL, and physical activity. For HbA_{1c}, short-term improvement was significant (standardized mean difference [SMD] –0.43; 95% CI –0.64 to –0.21; I^2=69%) and medium term (SMD –0.49; 95% CI –0.49 to –0.09; I^2=21%). However, in the long term, the improvement was not significant (SMD –0.23; 95% CI –0.49 to 0.03; I^2=88%). For QoL, significant improvements were observed in the short term (SMD –0.23; 95% CI –0.42 to –0.05; I^2=62%), and in the medium term (SMD –0.16; 95% CI –0.24 to –0.07; I^2=0%). In the long term, however, the improvement was not significant (SMD –0.12; 95% CI –0.41 to 0.16; I^2=71%). For physical activity, both subjective (questionnaire) and objective (number of steps) outcomes were analyzed. In the short term, subjective outcomes showed significant improvement (SMD 0.31; 95% CI 0.12-0.50; I^2=0%), while objective outcomes did not (SMD 0.11; 95% CI –0.05 to 0.27; I^2=0%). Medium- and long-term subjective outcomes showed no significant improvement. Meta-analysis for objective outcomes in the medium and long term was not possible due to insufficient studies.

Conclusions: This study confirmed short- and medium-term benefits of mHealth combined with professional interventions for HbA_{1c}, QoL, and short-term physical activity, supporting effective chronic disease management.

XSL•FO
RenderX

(Interact J Med Res 2025;14:e55835) doi:10.2196/55835

KEYWORDS

mHealth; systematic reviews; meta-analysis; chronic diseases; global health; technology; health care; interventions; chronic conditions; health care professionals; World Health Organization; physical activity; web-based

Introduction

Background

Mobile health (mHealth) is the delivery of medical and public health services using mobile handsets, mobile kiosks, and other wireless terminals in real-world environments [1]. Interventions in mHealth can take various forms, including the creation of original apps for the management of health conditions and communication, simple text message-based interventions, as well as those using voice calls or the accumulation of device data on mobile apps for data management [2]. Many interventions in the field of mHealth primarily emphasize self-management without any direct human involvement [3]. Likewise, there are interventions that involve periodic health care professional assistance and the additional use of health devices [4,5].

The growing disease burden of chronic diseases, on the other hand, is one of the global challenges [6]. Chronic diseases include conditions such as cardiovascular diseases, stroke, cancer, diabetes, respiratory conditions, and arthritis [7]. Furthermore, obesity is now defined as a disease, and the most common chronic disease [8]. While the assistance of health care professionals is essential for the management of these chronic conditions, mHealth may be valuable in filling the gaps between medical consultations or support, as well as in the monitoring of disease conditions remotely. In fact, the efficacy of therapeutic apps prescribed by physicians has also been reported [9]. The effectiveness of mHealth is currently being investigated through a systematic review [10]. The quality of research methodologies is often limited, and consistent conclusions have not been reached yet. Among the reasons for this is that the definition of mHealth intervention methods varies widely, and depending on the definition, conclusions about effectiveness can differ. Some reports suggest that no significant effects have been observed. There is a discrepancy in research on health interventions that focus on self-monitoring for chronic diseases, and it is unclear if there are any conclusive benefits [11-13]. A systematic review evaluating the effectiveness of remote monitoring in various types of chronic diseases found only limited improvements in general health status and clinical outcomes [14].

From a behavioral science perspective, effective management of people with chronic conditions requires more than just the use of mHealth technologies. It also requires consistent support from health care professionals over medium and long terms. In particular, studies focused on obesity have used interaction with professionals as an indicator of engagement. This interaction has been shown to play a significant role in achieving long-term (ie, 1 year) weight loss outcomes [15] although these definitions may vary between studies, potentially affecting the interpretation of results. Therefore, to maximize the potential of mHealth, we hypothesized that combining mHealth with health professional–led interventions would be more effective than relying solely on health care use. To investigate this hypothesis, our study focused exclusively on interventions that integrated mHealth with professional-led chronic disease strategies and then evaluated their effectiveness for each different duration.

Objectives

With this systematic review and meta-analysis, we investigated the effectiveness of combining mHealth and health professional–led intervention for improving health-related outcomes in chronic diseases.

Methods

Study Design

This was a systematic review and meta-analysis of randomized controlled trials.

Protocol and Registration

This systematic review was preregistered in the PROSPERO database (CRD42022337882) and conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1).

Search Strategy

Our search spanned 4 databases from their start dates until July 17, 2023. These included Web of Science, CENTRAL, MEDLINE, and CINAHL, as they are the four most used databases in similar research. Manual searches systematically screened the reference lists of all studies identified in the database search. The same inclusion criteria were applied to these references and the titles and abstracts were reviewed to determine whether they were included studies. Our systematic search strategy used both medical subject heading terms and keywords derived from subject headings. We limited our inclusion to randomized controlled trials published in English, excluding gray literature. The detailed search strategy is available in Multimedia Appendix 2.

Identification and Selection of Trials

Population

The population of interest included patients aged 18 years and older with at least 1 chronic condition. This was a change from the original protocol (CRD42022337882), which specified a minimum age of 20 years, in order to increase the generalizability of the findings by including all adults.

Interventions

This study's interventions comprised a mix of mHealth and direct health care professional involvement. The World Health Organization defines mHealth as "the use of mobile and wireless technologies to support health objectives" [16]. We excluded interventions that primarily used computers, categorizing them

XSL•FO
RenderX

as eHealth. Additionally, interventions where participants used smartphones solely for calls or text messages, without using specific apps, were not included to avoid overlapping mHealth and eHealth interventions and to ensure a clear distinction. The health care professionals involved in our study spanned various fields, including doctors, nurses, physiotherapists, occupational therapists, and dieticians. These professionals were actively engaged in delivering the mHealth intervention, beyond just explaining or reviewing app data. The methods used by health professionals for interventions included face-to-face interactions, phone calls, and text messaging.

Comparison

The comparison groups in this study consisted of the general control groups used in each individual study. For example, control groups were defined as those with only general care and follow-up, or those using mHealth-related devices but no involvement of a health care professional.

Outcomes

This systematic review assessed the following primary outcomes: HbA_{1c} levels, quality of life (QoL), and physical activity. These outcomes were chosen for their relevance in chronic disease management and were analyzed across different follow-up periods (short, medium, and long terms) to evaluate intervention effectiveness over time.

Selection Process

Rayyan software was used. Two authors (MK and TS) independently assessed the title and abstract records to identify eligible studies. The same authors then conducted a full-text screening to assess whether the criteria were met; in cases where the 2 authors' judgments conflicted, a third author (TM) decided and ultimately reached a consensus.

Assessment of Methodological Quality

The risk of bias for individual studies was assessed using the Cochrane Risk of Bias Framework for sequence generation, allocation concealment, blinding, completeness of outcome assessment, selective reporting, and other biases. Data were extracted by 2 authors (TM and MK). Disagreements were reviewed by a third author and a final consensus was reached. For the quality of the clinical evidence, the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) system was used to evaluate the results.

Data Extraction

Data were extracted by 2 authors (TM and MK). Disagreements were reviewed by a third reviewer and a final consensus was reached. Data were extracted on (1) basic information about the study, such as authors, year of publication, country, and region where the intervention took place; (2) target population; (3) intervention group description; (4) control group description; (5) outcomes; and (6) key findings.

Data Analysis

When two or more studies assessed the same outcome, a meta-analysis was performed using Review Manager 5.4 (Nordic Cochrane Centre). The results of the intervention groups were compared with the control group in each meta-analysis. The combined app and professional intervention was considered the intervention group and compared to the control group when there were more than 2 groups. The meta-analysis evaluated 3 different categories of time periods: short-, medium-, and long-term effects. Short term was defined as the closest to 3 months, medium term as the closest to 6 months, and long term as the closest to 12 months. For each follow-up point, the mean and SD were calculated. The group mean was calculated from the SE, CI, and *P* value of the difference between the means in trials that did not report SDs for each intervention group. In studies that only reported the median and IQR without providing SDs, we estimated the mean using the median value directly and approximated the SD by multiplying the IQR by 1.35. This method follows established meta-analytic conventions for handling summary statistics in studies with limited data [17]. All meta-analyses were presented as standardized mean difference (SMD) to account for adjusted and unadjusted means, due to the different data presentation formats of the included studies [17]. The authors of the studies in which the data were presented in a graphical format were contacted to obtain the exact values. If the reporting of results was still unclear after the above procedure, studies were excluded from the meta-analysis. For each meta-analysis, heterogeneity was assessed using I^2. When the I^2 value exceeded 75%, sensitivity analysis was conducted by sequentially excluding studies with the largest effect sizes, as these studies contributed significantly to the overall heterogeneity. This stepwise exclusion continued until heterogeneity dropped below the 75% threshold. By reducing the influence of studies with disproportionately high effect sizes, we aimed to create a more consistent and interpretable dataset, enhancing the reliability of the pooled effect size and minimizing the risk of bias or misinterpretation due to extreme variability among studies [18].

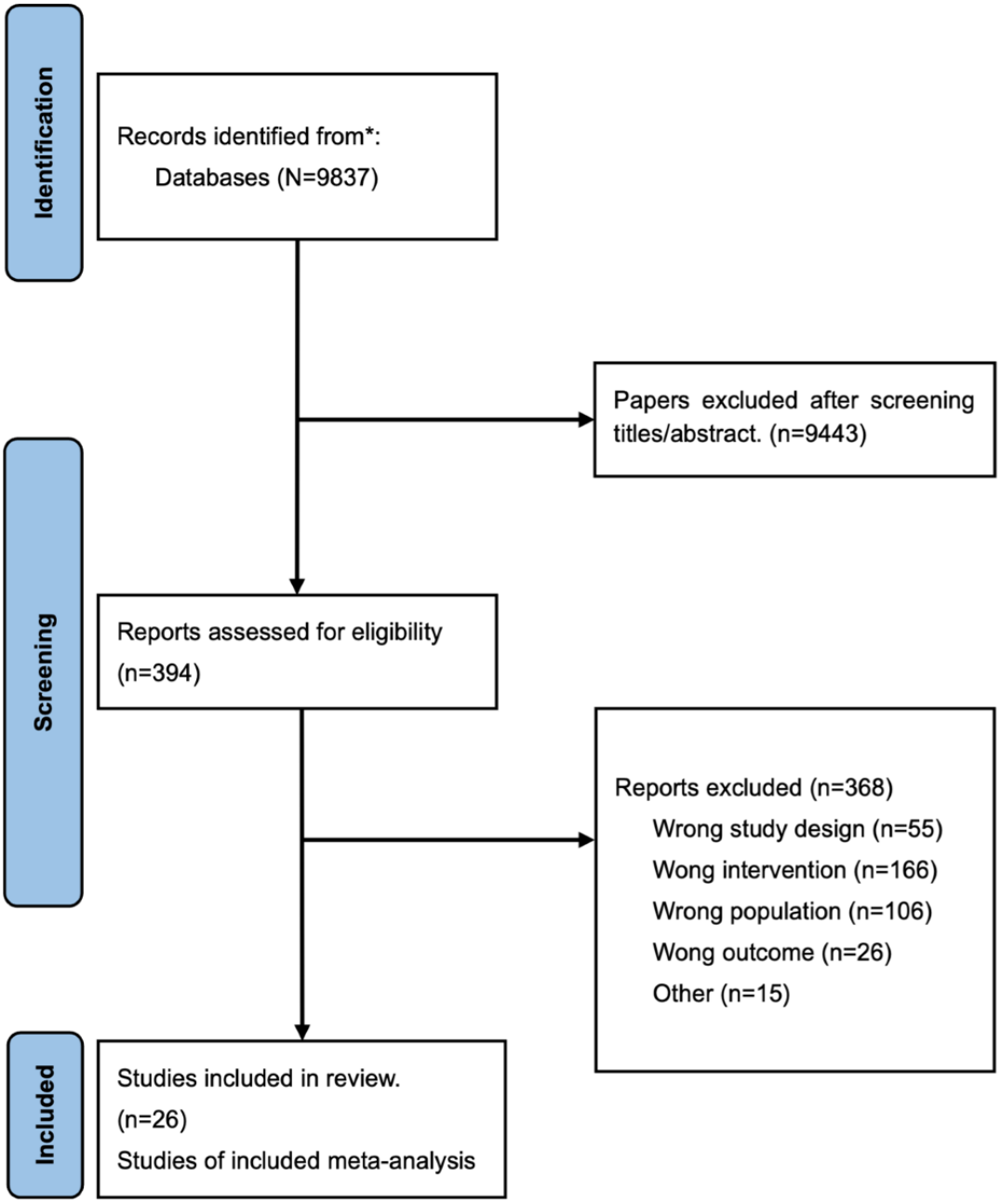
Results

Study Selection

The initial search identified 9837 records, which were screened for title and text as primary screening after removing duplicates. A number of 394 were selected for secondary screening after the first screening. Finally, 26 studies were finally selected. Figure 1 shows the flow chart of the selection process.

XSL•FO
RenderX

Figure 1. Flow diagram.



Characteristics of Included Trials and Intervention Description

Regarding the country, studies were conducted at outpatient or community practices in the following countries: Pakistan (2/26, 8%) [19,20], Spain (4/26, 15%) [21-24], Mexico (2/26, 8%) [25,26], United Kingdom (4/26, 15%) [27-30], United States (2/26, 8%) [31,32], China (4/26, 15%) [33-36], Singapore (2/26, 8%) [34,37]; and one each in Netherlands [38], Australia [39], Canada [40], Germany [41], Indonesia [42], and Thailand [43].

Regarding the disease, diabetes was the most common disease, with 12 of 26 trials [21,22,25,29-31,33,34,36,39,42,43]. Seven studies included type 2 diabetes [21,25,30,31,33,34,39] and 5 studies included both [22,29,36,42,43]. No study included only type 1 diabetes. The next most common was heart disease, which accounted for 7 of the 26 cases [19,20,24,35,37,38,44]. Other disease areas included cancer (2 studies) [26,32], respiratory diseases (1 study) [28], chronic diseases (2 studies) [23,40], metabolic syndrome (1 study) [41], and depression (1 study) [27].

The most common outcome was quality of life. There were 15 studies [20,22,24-27,30,32,34,37-41,44]. Ten studies used physical activity as an outcome [19,21,26,28,30,32,35,39,40,44]. Details are provided in Multimedia Appendix 3 [19-44].

Risk of Bias Assessment

The results of the risk of bias assessment are summarized in Multimedia Appendix 4 [19-44]. The risk of bias assessment for studies on HbA_{1c}, QoL, and physical activity (steps and subjective measures) showed that most studies had a low risk of bias. For example, studies like Anzaldo-Campos et al [25] and Blair et al [26] were evaluated as low-risk. Some studies had some concerns, such as Azelton et al [31] and Gill et al [40], but none were classified as high risk. Several studies, including

XSL•FO
RenderX

Franc et al [22] and Manzoor et al [19], did not have sufficient data for a full assessment.

Certainty of the Evidence According to the GRADE Approach

The certainty of evidence results using the GRADE approach are described in the results for each outcome. The GRADE assessment showed varying levels of evidence quality across outcomes. Most HbA_{1c} outcomes had very low to moderate certainty, with short-term outcomes showing very low certainty and medium-term outcomes showing moderate certainty. QoL outcomes generally had low to moderate certainty, with long-term outcomes showing very low certainty. Physical activity outcomes varied, with step counts showing moderate certainty and subjective measures ranging from very low to moderate certainty. These findings highlight the need for caution due to risks of bias and inconsistencies. A table with the results is provided in Multimedia Appendix 5.

Synthesis of Results: Meta-Analysis

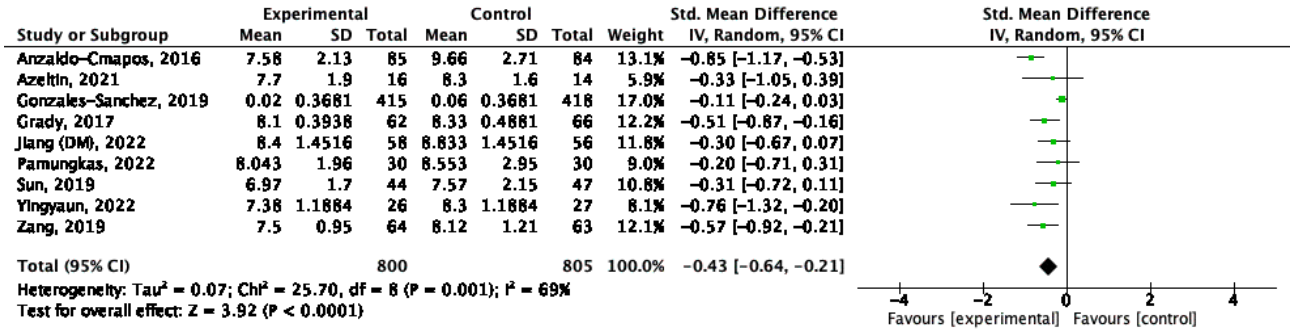
One study was excluded because data were not available [24]. Therefore, a meta-analysis was performed from 25 studies. Our systematic review included a broad spectrum of conditions, but diabetes was the most frequently addressed condition among the included studies. Consequently, HbA_{1c} was chosen as a primary indicator due to its importance in diabetes management and its frequent reporting in the studies. QoL was selected as it is a critical outcome measure reflecting the overall well-being of patients with chronic conditions. Physical activity and subjective outcomes of physical activity were included because these measures are commonly used to assess lifestyle modifications and their impact on chronic disease management. These indicators were the most frequently reported across the studies we reviewed, making them the most relevant for our analysis.

HbA_{1c} Outcome

Short Term

Meta-analysis was performed on 9 studies [23,25,29,31,33,34,36,42,43]. There was a significant increase in improvement for the combination intervention (SMD –0.43; 95% CI –0.64 to –0.21; P<.001; I^2=69%; Figure 2 [23,25,29,31,33,34,36,42,43]).

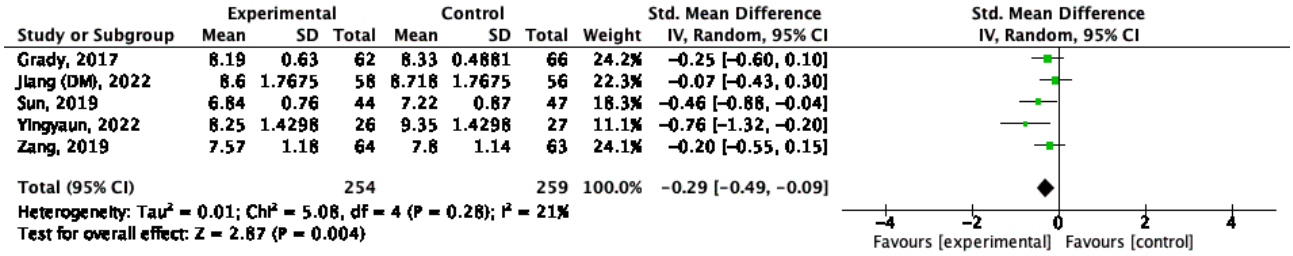
Figure 2. Hemoglobin A1c (HbA_{1c}) for short term [23,25,29,31,33,34,36,42,43].



Medium Term

A meta-analysis was performed on 5 studies [29,33,34,36,43]. There was a significant increase in improvement for the combination intervention (SMD –0.49; 95% CI –0.49 to –0.09; P<.001; I^2=21%; Figure 3 [29,33,34,36,43]).

Figure 3. Hemoglobin A1c (HbA_{1c}) for medium term [29,33,34,36,43].

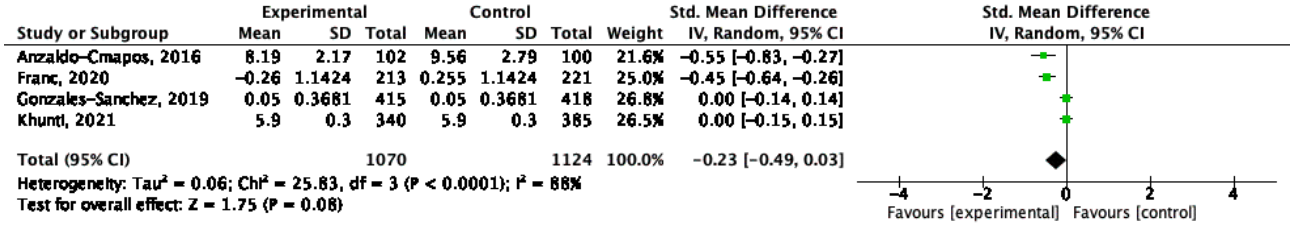


Long Term

One study was excluded because data were not available [24]. A meta-analysis was performed on 4 studies [22,23,25,30]. There was no significant increase in improvement for the combination intervention in the long term (SMD –0.23; 95% CI –0.49 to 0.03; P<.001; I^2=88%; Figure 4 [22,23,25,30]). Sensitivity analysis was performed because heterogeneity was >75%. Two studies [23,30] were excluded and did not reach statistical significance (SMD –0.00; 95% CI –0.10 to 0.10; P<.001; I^2=0%).

XSL•FO
RenderX

Figure 4. Hemoglobin A1c (HbA_{1c}) for long term [22,23,25,30].

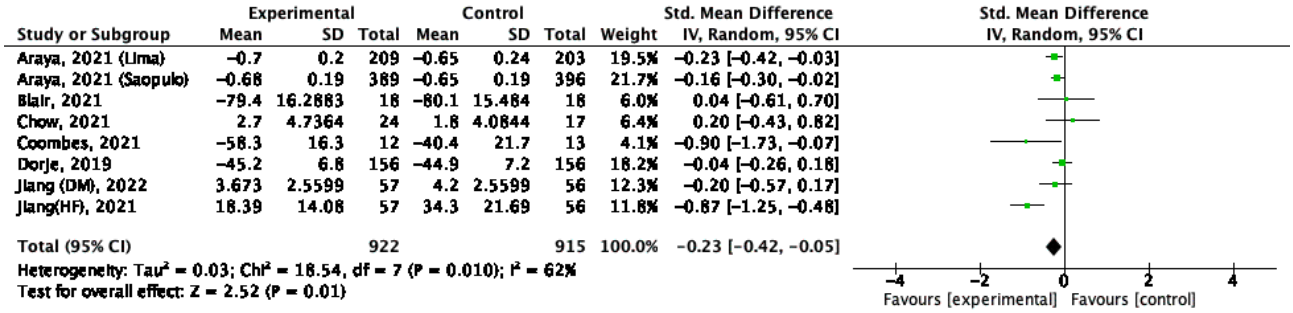


QoL Outcome

Short Term

Seven studies were included. However, 1 study [27] considered each of the 2 real samples, so they were included in the meta-analysis separately. Therefore, a meta-analysis was performed as 8 studies [26,27,32,34,37,39,44]. There was a significant increase in improvement for the combination intervention in the short term (SMD –0.23; 95% CI –0.42 to –0.05; P<.001; I^2=62%; Figure 5 [26,27,32,34,37,39,44]).

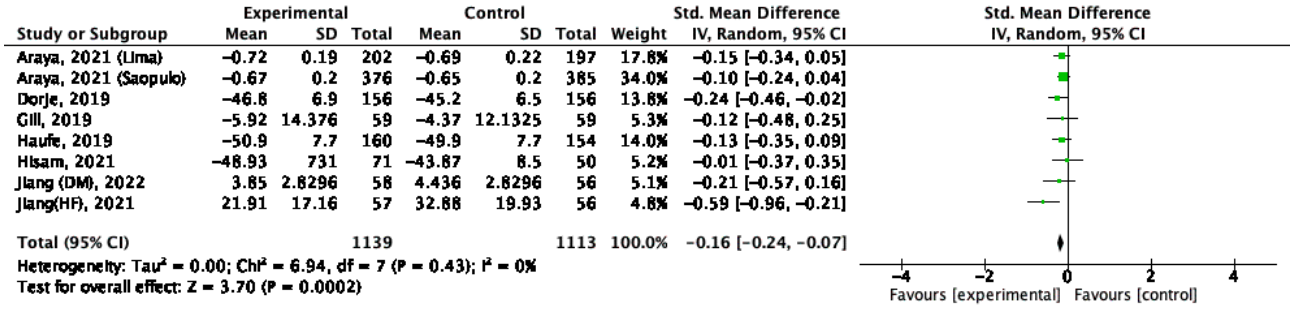
Figure 5. Quality of life for short term [26,27,32,34,37,39,44].



Medium Term

Seven studies were included. However, 1 study [27] considered each of the 2 real samples, so they were included in the meta-analysis separately. Therefore, a meta-analysis was performed as 8 studies [20,27,34,37,40,41,44]. There was a significant increase in improvement for the combination intervention (SMD –0.16; 95% CI –0.24 to –0.07; P<.001; I^2=0%; Figure 6 [20,27,34,37,40,41,44]).

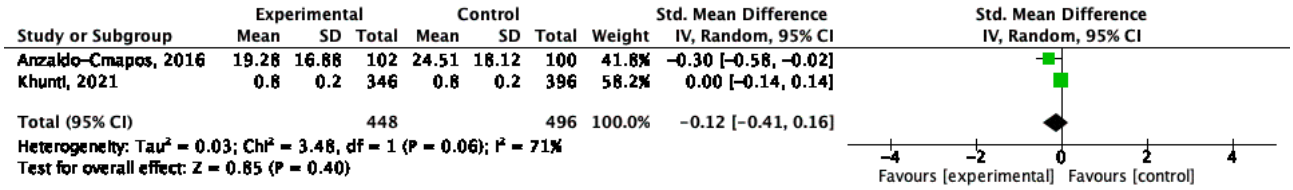
Figure 6. Quality of life for medium term [20,27,34,37,40,41,44].



Long Term

One study was excluded because data was not available [24]. A meta-analysis was performed on 2 studies [25,30]. There was no significant increase in improvement for the combination intervention (SMD –0.12; 95% CI –0.41 to 0.16; P=.04; I^2=71%; Figure 7 [25,30]).

Figure 7. Quality of life for long term [25,30].



XSL•FO
RenderX

Physical Activity

We performed a meta-analysis of the outcome of physical activity, dividing the data into subjective (questionnaire) and objective (number of steps).

Objective Outcomes of Physical Activity

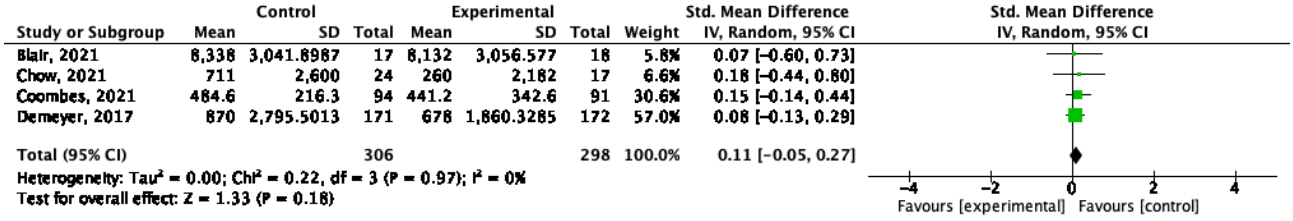
Overview

It was not possible to perform a meta-analysis for the medium and long terms because there was less than 1 study that corresponded to this time.

Short Term

Meta-analysis was conducted in 4 studies [26,28,32,39]. There was no significant increase in improvement for the combination intervention (SMD 0.11; 95% CI –0.05 to 0.27; P=.18; I^2=0%; Figure 8 [26,28,32,39]).

Figure 8. Physical activity (steps) for short term [26,28,32,39].

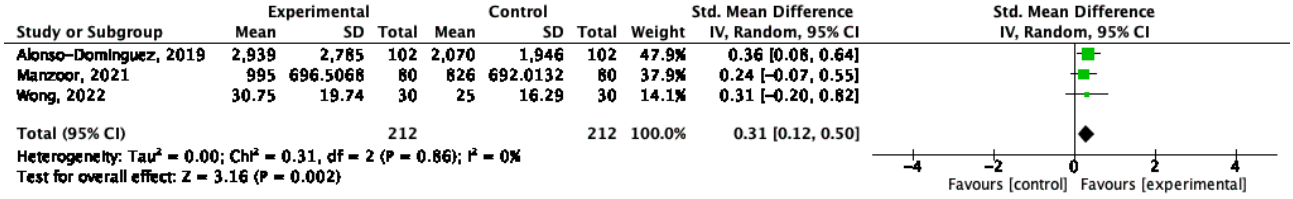


Subjective Outcomes of Physical Activity

Short Term

A meta-analysis of 3 studies was conducted [19,21,35]. There was a significant increase in improvement for the combination intervention (SMD 0.31; 95% CI 0.12-0.50; P=.002; I^2=0%; Figure 9 [19,21,35]).

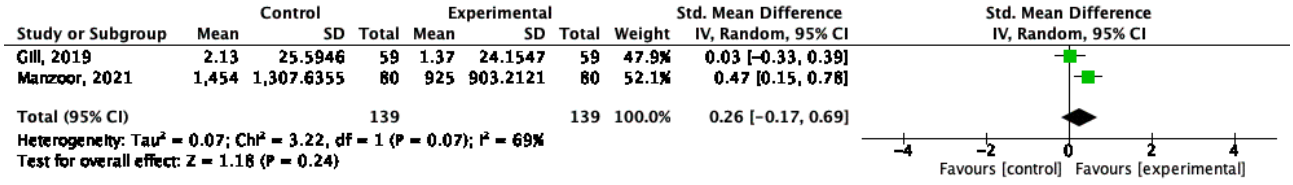
Figure 9. Physical activity (subjective) for short term [19,21,35].



Medium Term

Meta-analysis was conducted in 2 studies [19,40]. There was no significant increase in improvement for the combination intervention (SMD 0.26; 95% CI –0.17 to 0.69; P=.24; I^2=69%; Figure 10 [19,40]).

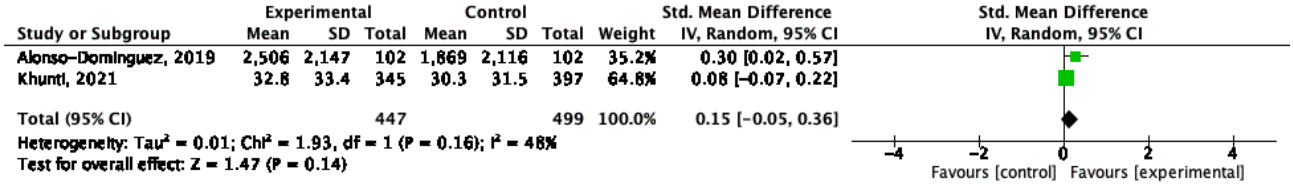
Figure 10. Physical activity (subjective) for medium term [19,40].



Long Term

Meta-analysis was conducted in 2 studies [21,30]. There was no significant increase in improvement for the combination intervention (SMD 0.15; 95% CI –0.05 to 0.36; P=.14; I^2=48%; Figure 11 [21,30]).

Figure 11. Physical activity (subjective) for long term [21,30].



XSL•FO
RenderX

Discussion

Principal Findings

This systematic review and meta-analysis showed that the combination of mHealth and health professional–led interventions improved health-related outcomes such as HbA_{1c} and QoL in chronic diseases. These effects were significant in the short and medium terms and there was no significant long term. For physical activity, there were no differences in intervention effects except for the short term as measured by questionnaires.

HbA_{1c}

In this systematic review, studies targeting patients with chronic diseases were included, and among the 26 studies, 12 incorporated HbA_{1c} as an outcome measure. In a meta-analysis that used mobile apps for lifestyle modification among diabetes patients, similar to our study, improvements in HbA_{1c} levels were demonstrated [45]. Furthermore, the meta-analysis indicated significant long-term effects in its subgroup analysis, defining long-term as 9-12 months [45], which differs from our study's definition. The impact of mHealth was indicated to diminish in the long term due to factors such as adherence fatigue [46]. However, another meta-analysis found that digital health–led diabetes self-management education and support (including mHealth, eHealth, and interventions using social networking services) were effective in improving HbA_{1c} at 6 and 12 months [47]. The discrepancies between our study and the other meta-analysis may be due to some factors. These factors include variations in digital health interventions, professional involvement levels, patient populations, and adherence. The other meta-analysis also included mHealth, eHealth, and interventions using social networking services, which may offer different engagement and support levels. These differences in intervention design and implementation could contribute to the varying long-term effectiveness observed. An overview of the studies we included in our systematic review, for example, Anzaldo-Campos et al [25] demonstrated that a smartphone app used for diabetes management significantly improved HbA_{1c} levels over a 6-month period, highlighting the potential of app-based interventions for glycemic control. In contrast, Franc et al [22] found that a telemonitoring system improved HbA_{1c} levels significantly over 12 months, suggesting that the integration of professional support enhances the effectiveness of digital health interventions. The differing results could be due to the varying levels of professional involvement and continuous monitoring provided in the telemonitoring system, which may help sustain patient adherence and engagement over a longer period. Additionally, Jiang et al [37] found that a nurse-led smartphone intervention significantly reduced HbA_{1c} levels over 6 months in patients with type 2 diabetes, emphasizing the importance of professional guidance in mHealth interventions. Conversely, Zhang et al [36] reported that a self-guided mobile app intervention without professional support did not result in significant improvements in HbA_{1c} levels, highlighting the potential limitations of mHealth interventions lacking continuous professional involvement. Our findings suggest that while the combination of mHealth and health professional–led interventions can be as diverse as those described above, continuous feedback and support to patients through other digital interventions to enhance engagement may be a key factor in achieving long-term benefits [48].

QoL

QoL was the most prevalent outcome measure among the studies included in this study. The results of our meta-analysis suggested that combining mHealth and health professional–led intervention can significantly improve QoL in chronic diseases in the short and medium terms. Studies conducting meta-analyses on the QoL in mHealth interventions for patients with chronic diseases are limited [49-51]. In a meta-analysis by Qin et al [50], mobile app–based interventions improved the QoL (SMD 0.39; 95% CI 0.27-0.51; $P<.001$) of patients with cancer. They emphasized that short-term interventions in particular (duration of 3 months or less), physician-patient interaction interventions, and cognitive behavioral therapy–based interventions may be most effective in improving QoL, which was consistent with this study. While extending beyond the scope of mHealth interventions, there is only 1 reported meta-analysis that indicated long-term improvements in QoL (mean difference 0.92; 95% CI 0.06-1.78; $P=.04$) resulting from cardiac telerehabilitation for patients with coronary artery disease [52]. For instance, Araya et al [27] found that a digital intervention combined with professional support significantly improved QoL in patients with cardiovascular diseases over 12 months. This improvement could be attributed to the continuous monitoring and personalized feedback provided by health care professionals, which likely enhanced patient adherence and engagement with the intervention. In contrast, Blasco et al [24] reported improvements in QoL among patients with acute coronary syndrome using a web-based telemonitoring system. The significant improvements observed in Blasco et al [24] study might be due to the comprehensive nature of the telemonitoring system, which included regular remote check-ins and the ability to promptly address patient concerns, thereby providing a sense of security and continuous care. Similarly, Manzoor et al [19] demonstrated that the mHealth-augmented cardiac rehabilitation program significantly improved QoL in patients with postacute coronary syndrome over 24 weeks, underscoring the importance of combining digital interventions with structured professional support. On the other hand, Wong et al [35] found that a self-guided mHealth intervention without professional involvement had limited effects on improving QoL, indicating that professional support is crucial for maximizing the benefits of digital health interventions. In our meta-analysis, focusing solely on mHealth interventions, it might be possible to identify long-term effects on the improvement of QoL by expanding the scope to include telehealth intervention or telerehabilitation as well.

Physical Activity

The results of our meta-analysis suggested that physical activity had no significant effect except for the short-term effect of subjective assessment. One reason for this might be that many mHealth interventions did not include specifications or support

XSL•FO
RenderX

to promote physical activity. This could be attributed to the predominant focus of many interventions on monitoring blood data and providing advice [22,24,25,29,31,33,34,36,42,43], potentially falling short of implementing behavior changes that could lead to promoting physical activity. In fact, among the 26 studies, only 8 studies adopted physical activity as the primary outcome. As an example of a study that adopted physical activity as the primary outcome, Gill et al [40] examined the effects of using a mobile app focused primarily on increasing physical activity and professional support for participants at risk for chronic disease and reported not only an increase in steps over the next 6 months but also maintenance of steps after 1 year. Dorje et al [44] found that a mHealth intervention including exercise modules improved physical activity levels in patients with coronary heart disease over 12 months, highlighting the potential of structured digital interventions. Conversely, Demeyer et al [28] reported that a self-guided digital intervention without professional support did not significantly increase physical activity levels in patients with chronic obstructive pulmonary disease, suggesting the need for continuous professional involvement to achieve sustained behavioral changes. In a different disease from the above, Blair et al [26] showed that a combination of mHealth and professional-led interventions for patients with cancer increased the number of steps taken, although there was no effect on outcomes related to sedentary behavior among physical activity measures. Thus, it would be valuable to identify in future studies which chronic diseases are more likely to improve physical activity with a combination of mHealth and professional-led interventions. Stavric et al [53] showed that digital physical activity and exercise interventions had positive effects on self-reported physical activity in people with chronic diseases, but not on objectively measured physical activity, which supports the results of our study. Their meta-analysis included self-guided digital mHealth interventions [53]. Therefore, there is room for future research to investigate whether self-directed digital or mHealth interventions or a combination of mHealth and health professional–led interventions, have more positive effects on physical activity.

Implications for Practice

The results of our study suggest that combining mHealth with professionally led interventions can be effective in chronic disease management, particularly in short and medium terms. This combination approach shows potential for improving important health outcomes such as HbA_{1c} and QoL. However, the lack of long-term effects indicates the need for sustained and adaptive intervention strategies. Health care providers should consider integrating ongoing digital support and feedback mechanisms to maintain patient engagement and adherence over time. In addition, the involvement of health care professionals in mHealth interventions should be designed to provide personalized and digital support, leveraging technologies such as telemedicine and mobile apps. For example, health care organizations should invest in training and infrastructure to support the seamless integration of mHealth technologies into traditional care models. This can improve the scalability and effectiveness of chronic disease management programs, ultimately leading to better patient outcomes and reduced health care costs.

Limitations

First, a major limitation of our study is the ambiguity in the definition of "health professional–led intervention." Our research focused on a combination of such interventions with mHealth strategies. However, the term may be misleading, as many studies typically include some level of health professional involvement in routine care. In contrast, our study specifically included cases with a more active role for health professionals than is typically seen in routine care. Additionally, the definition of the control group varied across studies. Some studies included only general care, while others provided mHealth devices without additional interventions. Second, our focus on health-related outcomes in chronic diseases limited our ability to conduct comprehensive meta-analyses due to the diverse nature of these diseases and their outcomes. Finally, we did not assess the cost-effectiveness of integrating health professional–led interventions with mHealth strategies. While the involvement of health professionals could potentially improve intervention effectiveness and reduce dropout rates, it could also lead to higher costs. Although mHealth interventions are known to struggle with high dropout rates [54], the cost implications of incorporating health professional support need to be further explored in future studies.

Conclusions

This systematic review and meta-analysis showed that the combined mHealth and health professional–led intervention had positive effects on HbA_{1c} and QoL in chronic conditions in short and medium terms. Positive effects on physical activity were only observed in the short term as measured by questionnaires. To achieve long-term effects of more than 1 year, it may be necessary to implement digital interventions that provide continuous feedback and support, tailored to a combination of mHealth and health professional–led interventions. Future research should explore the sustainability of these interventions over extended periods and investigate the specific components of mHealth and health professional input that contribute most to positive outcomes. Additionally, the integration of advanced technologies such as artificial intelligence and machine learning could further enhance the personalization and effectiveness of mHealth interventions.

Acknowledgments

The authors are grateful to Kono and Nemoto of Hokkaido University Library for their assistance in the development of the search formula.

XSL•FO
RenderX

Conflicts of Interest

PREVENT Inc is a company providing mHealth-based disease management programs. MK has received consulting fees from PREVENT Inc and is a nonregular staff member. TM and TS are employees of PREVENT Inc. YH is a founder and stockholder of PREVENT Inc.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 32 KB - ijmr_v14i1e55835_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 30 KB - ijmr_v14i1e55835_app2.docx]

Multimedia Appendix 3
Characteristics of included studies with author names beginning with A to Z.
[DOCX File , 32 KB - ijmr_v14i1e55835_app3.docx]

Multimedia Appendix 4
Risk of bias assessment.
[DOCX File , 66 KB - ijmr_v14i1e55835_app4.docx]

Multimedia Appendix 5
Certainty of the evidence according to the GRADE (Grading of Recommendations Assessment, Development, and Evaluation) approach.
[DOCX File , 20 KB - ijmr_v14i1e55835_app5.docx]

References

1. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010;15:1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
2. WHO Global Observatory for eHealth. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. Geneva, Switzerland: World Health Organization; 2011.
3. Kassavou A, Wang M, Mirzaei V, Shpendi S, Hasan R. The association between smartphone app-based self-monitoring of hypertension-related behaviors and reductions in high blood pressure: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10(7):e34767 [FREE Full text] [doi: 10.2196/34767] [Medline: 35819830]
4. Bollyky JB, Bravata D, Yang J, Williamson M, Schneider J. Remote lifestyle coaching plus a connected glucose meter with certified diabetes educator support improves glucose and weight loss for people with type 2 diabetes. J Diabetes Res 2018;2018:3961730 [FREE Full text] [doi: 10.1155/2018/3961730] [Medline: 29888288]
5. Kanai M, Toda T, Yamamoto K, Akimoto M, Hagiwara Y. A mobile health–based disease management program improves blood pressure in people with multiple lifestyle-related diseases at risk of developing vascular disease—a retrospective observational study. Circ Rep 2022;4(7):322-329 [FREE Full text] [doi: 10.1253/circrep.CR-22-0024] [Medline: 35860354]
6. Hajat C, Stein E. The global burden of multiple chronic conditions: a narrative review. Prev Med Rep 2018;12:284-293 [FREE Full text] [doi: 10.1016/j.pmedr.2018.10.008] [Medline: 30406006]
7. Schmidt H. Chronic disease prevention and health promotion. In: Barrett DH, Ortmann LW, Dawson A, Saenz C, Reis R, Bolan G, editors. Public Health Ethics: Cases Spanning the Globe. Cham, Switzerland: Springer; 2016.
8. Tucker S, Bramante C, Conroy M, Fitch A, Gilden A, Wittleder S, et al. The most undertreated chronic disease: addressing obesity in primary care settings. Curr Obes Rep 2021;10(3):396-408 [FREE Full text] [doi: 10.1007/s13679-021-00444-y] [Medline: 34297343]
9. Kario K, Nomura A, Harada N, Okura A, Nakagawa K, Tanigawa T, et al. Efficacy of a digital therapeutics system in the management of essential hypertension: the HERB-DH1 pivotal trial. Eur Heart J 2021;42(40):4111-4122 [FREE Full text] [doi: 10.1093/eurheartj/ehab559] [Medline: 34455443]
10. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
11. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]

XSL•FO
RenderX

12. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
13. Yi JY, Kim Y, Cho YM, Kim H. Self-management of chronic conditions using mHealth interventions in Korea: a systematic review. Healthc Inform Res 2018;24(3):187-197 [FREE Full text] [doi: 10.4258/hir.2018.24.3.187] [Medline: 30109152]
14. Castelyn G, Laranjo L, Schreier G, Gallego B. Predictive performance and impact of algorithms in remote monitoring of chronic conditions: a systematic review and meta-analysis. Int J Med Inform 2021;156:104620 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104620] [Medline: 34700194]
15. Carey A, Yang Q, DeLuca L, Toro-Ramos T, Kim Y, Michaelides A. The relationship between weight loss outcomes and engagement in a mobile behavioral change intervention: retrospective analysis. JMIR Mhealth Uhealth 2021;9(11):e30622 [FREE Full text] [doi: 10.2196/30622] [Medline: 34747706]
16. mHealth: Use of appropriate digital technologies for public health. World Health Organization. 2018. URL: https://apps.who.int/gb/ebwha/pdf_files/wha71/a71_20-en.pdf [accessed 2024-12-17]
17. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, NJ: John Wiley & Sons; 2019.
18. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
19. Manzoor S, Hisam A, Aziz S, Mashhadi SF, Haq ZU. Effectiveness of mobile health augmented cardiac rehabilitation on behavioural outcomes among post-acute coronary syndrome patients: a randomised controlled trial. J Coll Physicians Surg Pak 2021;31(10):1148-1153. [doi: 10.29271/jcpsp.2021.10.1148] [Medline: 34601832]
20. Hisam A, Haq ZU, Aziz S, Doherty P, Pell J. Effectiveness of mobile health augmented cardiac rehabilitation (MCard) on health-related quality of life among post-acute coronary syndrome patients: a randomized controlled trial. Pak J Med Sci 2022;38:716-723 [FREE Full text] [doi: 10.12669/pjms.38.3.4724] [Medline: 35480536]
21. Alonso-Domínguez R, Patino-Alonso MC, Sánchez-Aguadero N, García-Ortiz L, Recio-Rodríguez JI, Gómez-Marcos MA. Effect of a multifactorial intervention on the increase in physical activity in subjects with type 2 diabetes mellitus: a randomized clinical trial (EMID study). Eur J Cardiovasc Nurs 2019;18(5):399-409. [doi: 10.1177/1474515119835048] [Medline: 30808196]
22. Franc S, Hanaire H, Benhamou PY, Schaepelynck P, Catargi B, Farret A, et al. DIABEO system combining a mobile app software with and without telemonitoring versus standard care: a randomized controlled trial in diabetes patients poorly controlled with a basal-bolus insulin regimen. Diabetes Technol Ther 2020;22(12):904-911 [FREE Full text] [doi: 10.1089/dia.2020.0021] [Medline: 32407148]
23. Gonzalez-Sanchez J, Recio-Rodriguez JI, Fernandez-delRio A, Sanchez-Perez A, Magdalena-Belio JF, Gomez-Marcos MA, et al. Using a smartphone app in changing cardiovascular risk factors: a randomized controlled trial (EVIDENT II study). Int J Med Inform 2019;125:13-21. [doi: 10.1016/j.ijmedinf.2019.02.004] [Medline: 30914176]
24. Blasco A, Carmona M, Fernández-Lozano I, Salvador C, Pascual M, Sagredo P, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. J Cardiopulm Rehabil Prev 2012;32(1):25-31. [doi: 10.1097/HCR.0b013e3182343aa7] [Medline: 22113368]
25. Anzaldo-Campos MC, Contreras S, Vargas-Ojeda A, Menchaca-Díaz R, Fortmann A, Philis-Tsimikas A. Dulce wireless Tijuana: a randomized control trial evaluating the impact of project Dulce and short-term mobile technology on glycemic control in a family medicine clinic in Northern Mexico. Diabetes Technol Ther 2016;18(4):240-251 [FREE Full text] [doi: 10.1089/dia.2015.0283] [Medline: 26914371]
26. Blair CK, Harding E, Wiggins C, Kang H, Schwartz M, Tarnower A, et al. A home-based mobile health intervention to replace sedentary time with light physical activity in older cancer survivors: randomized controlled pilot trial. JMIR Cancer 2021;7(2):e18819 [FREE Full text] [doi: 10.2196/18819] [Medline: 33847588]
27. Araya R, Menezes PR, Claro HG, Brandt LR, Daley KL, Quayle J, et al. Effect of a digital intervention on depressive symptoms in patients with comorbid hypertension or diabetes in Brazil and Peru: two randomized clinical trials. JAMA 2021;325(18):1852-1862 [FREE Full text] [doi: 10.1001/jama.2021.4348] [Medline: 33974019]
28. Demeyer H, Louvaris Z, Frei A, Rabinovich RA, de Jong C, Gimeno-Santos E, et al. Physical activity is increased by a 12-week semiautomated telecoaching programme in patients with COPD: a multicentre randomised controlled trial. Thorax 2017;72(5):415-423 [FREE Full text] [doi: 10.1136/thoraxjnl-2016-209026] [Medline: 28137918]
29. Grady M, Katz LB, Cameron H, Levy BL. Diabetes app-related text messages from health care professionals in conjunction with a new wireless glucose meter with a color range indicator improves glycemic control in patients with type 1 and type 2 diabetes: randomized controlled trial. JMIR Diabetes 2017;2(2):e19 [FREE Full text] [doi: 10.2196/diabetes.7454] [Medline: 30291092]
30. Khunti K, Griffin S, Brennan A, Dallosso H, Davies M, Eborall H, et al. Behavioural interventions to promote physical activity in a multiethnic population at high risk of diabetes: PROPELS three-arm RCT. Health Technol Assess 2021;25(77):1-190 [FREE Full text] [doi: 10.3310/hta25770] [Medline: 34995176]

XSL•FO
RenderX

31. Azelton KR, Crowley AP, Vence N, Underwood K, Morris G, Kelly J, et al. Digital health coaching for type 2 diabetes: randomized controlled trial of healthy at home. Front Digit Health 2021;3:764735 [FREE Full text] [doi: 10.3389/fdgth.2021.764735] [Medline: 34901926]
32. Chow EJ, Doody DR, Di C, Armenian SH, Baker KS, Bricker JB, et al. Feasibility of a behavioral intervention using mobile health applications to reduce cardiovascular risk factors in cancer survivors: a pilot randomized controlled trial. J Cancer Surviv 2021;15(4):554-563 [FREE Full text] [doi: 10.1007/s11764-020-00949-w] [Medline: 33037989]
33. Sun C, Sun L, Xi S, Zhang H, Wang H, Feng Y, et al. Mobile phone–based telemedicine practice in older Chinese patients with type 2 diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2019;7(1):e10664 [FREE Full text] [doi: 10.2196/10664] [Medline: 30609983]
34. Jiang Y, Ramachandran HJ, Teo JYC, Leong FL, Lim ST, Nguyen HD, et al. Effectiveness of a nurse-led smartphone-based self-management programme for people with poorly controlled type 2 diabetes: a randomized controlled trial. J Adv Nurs 2022;78(4):1154-1165. [doi: 10.1111/jan.15178] [Medline: 35170786]
35. Wong EML, Leung DYP, Tam HL, Ko SY, Leung AYM, Lam SC, et al. Effectiveness of a nurse-led support programme using a mobile application versus phone advice on patients at risk of coronary heart disease - a pilot randomized controlled trial. Risk Manag Healthc Policy 2022;15:597-610 [FREE Full text] [doi: 10.2147/RMHP.S355554] [Medline: 35422666]
36. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of smartphone app–based interactive management on glycemic control in Chinese patients with poorly controlled diabetes: randomized controlled trial. J Med Internet Res 2019;21(12):e15401 [FREE Full text] [doi: 10.2196/15401] [Medline: 31815677]
37. Jiang Y, Koh KWL, Ramachandran HJ, Nguyen HD, Lim DS, Tay YK, et al. The effectiveness of a nurse-led home-based heart failure self-management programme (the HOM-HEMP) for patients with chronic heart failure: a three-arm stratified randomized controlled trial. Int J Nurs Stud 2021;122:104026. [doi: 10.1016/j.ijnurstu.2021.104026] [Medline: 34271265]
38. Broers ER, Widdershoven J, Denollet J, Lodder P, Kop WJ, Wetzels M, et al. Personalized eHealth Program for life-style change: results from the "Do Cardiac Health Advanced New Generated Ecosystem (Do CHANGE 2)" randomized controlled trial. Psychosom Med 2020;82(4):409-419. [doi: 10.1097/PSY.0000000000000802] [Medline: 32176191]
39. Coombes JS, Keating SE, Mielke GI, Fassett RG, Coombes BK, O'Leary KP, et al. Personal activity intelligence e-Health program in people with type 2 diabetes: a pilot randomized controlled trial. Med Sci Sports Exerc 2022;54(1):18-27. [doi: 10.1249/MSS.0000000000002768] [Medline: 34334715]
40. Gill DP, Blunt W, Boa Sorte Silva NC, Stiller-Moldovan C, Zou GY, Petrella RJ. The HealtheSteps™ lifestyle prescription program to improve physical activity and modifiable risk factors for chronic disease: a pragmatic randomized controlled trial. BMC Public Health 2019;19(1):841 [FREE Full text] [doi: 10.1186/s12889-019-7141-2] [Medline: 31253112]
41. Haufe S, Kerling A, Protte G, Bayerle P, Stenner HT, Rolff S, et al. Telemonitoring-supported exercise training, metabolic syndrome severity, and work ability in company employees: a randomised controlled trial. Lancet Public Health 2019;4(7):e343-e352 [FREE Full text] [doi: 10.1016/S2468-2667(19)30075-1] [Medline: 31204284]
42. Pamungkas RA, Usman AM, Chamroonsawasdi K, Abdurrasyid. A smartphone application of diabetes coaching intervention to prevent the onset of complications and to improve diabetes self-management: a randomized control trial. Diabetes Metab Syndr 2022;16(7):102537. [doi: 10.1016/j.dsx.2022.102537] [Medline: 35724489]
43. Yingyaun K, Charoensri S, Pongchaiyakul C. Effect of using mobile messenger for insulin injection education for glycemic control: a randomized controlled trial. Pac Rim Int J Nurs Res 2022;26:461-473 [FREE Full text]
44. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019;1(7):e363-e374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]
45. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
46. Lee JJN, Abdul Aziz A, Chan S, Raja Abdul Sahrizan RSFB, Ooi AYY, Teh YT, et al. Effects of mobile health interventions on health-related outcomes in older adults with type 2 diabetes: a systematic review and meta-analysis. J Diabetes 2023;15(1):47-57 [FREE Full text] [doi: 10.1111/1753-0407.13346] [Medline: 36649940]
47. Nkhoma DE, Soko CJ, Bowrin P, Manga YB, Greenfield D, Househ M, et al. Digital interventions self-management education for type 1 and 2 diabetes: a systematic review and meta-analysis. Comput Methods Programs Biomed 2021;210:106370. [doi: 10.1016/j.cmpb.2021.106370] [Medline: 34492544]
48. Sepah SC, Jiang L, Ellis RJ, McDermott K, Peters AL. Engagement and outcomes in a digital Diabetes Prevention Program: 3-year update. BMJ Open Diabetes Res Care 2017;5(1):e000422 [FREE Full text] [doi: 10.1136/bmjdrc-2017-000422] [Medline: 28948027]
49. Chin-Jung L, Hsiao-Yean C, Yeu-Hui C, Kuan-Chia L, Hui-Chuan H. Effects of mobile health interventions on improving glycemic stability and quality of life in patients with type 1 diabetes: a meta-analysis. Res Nurs Health 2021;44(1):187-200. [doi: 10.1002/nur.22094] [Medline: 33368403]
50. Qin M, Chen B, Sun S, Liu X. Effect of mobile phone app–based interventions on quality of life and psychological symptoms among adult cancer survivors: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2022;24(12):e39799 [FREE Full text] [doi: 10.2196/39799] [Medline: 36534460]

XSL•FO
RenderX

51. Gong Y, Jiang X, Chen X, Chen S, Wen Y, Yuan X, et al. Effectiveness of mHealth diet interventions in cancer survivors: a systematic review and meta-analysis of randomized controlled trials. Asia Pac J Oncol Nurs 2023;10(3):100196 [FREE Full text] [doi: 10.1016/j.apjon.2023.100196] [Medline: 37124242]
52. Zhong W, Liu R, Cheng H, Xu L, Wang L, He C, et al. Longer-term effects of cardiac telerehabilitation on patients with coronary artery disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2023;11:e46359 [FREE Full text] [doi: 10.2196/46359] [Medline: 37505803]
53. Stavric V, Kayes NM, Rashid U, Saywell NL. The effectiveness of self-guided digital interventions to improve physical activity and exercise outcomes for people with chronic conditions: a systematic review and meta-analysis. Front Rehabil Sci 2022;3:925620 [FREE Full text] [doi: 10.3389/fresc.2022.925620] [Medline: 36188933]
54. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. J Med Internet Res 2020;22(9):e20283 [FREE Full text] [doi: 10.2196/20283] [Medline: 32990635]

Abbreviations

GRADE: Grading of Recommendations Assessment, Development, and Evaluation
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-analyses
QoL: quality of life
SMD: standardized mean difference

Edited by T de Azevedo Cardoso; submitted 27.12.23; peer-reviewed by L Benning, M Fraser; comments to author 23.05.24; revised version received 19.06.24; accepted 03.12.24; published 20.01.25.

Please cite as:
Kanai M, Miki T, Sakoda T, Hagiwara Y
The Effect of Combining mHealth and Health Professional–Led Intervention for Improving Health-Related Outcomes in Chronic Diseases: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e55835
URL: https://www.i-jmr.org/2025/1/e55835
doi:10.2196/55835
PMID:

©Masashi Kanai, Takahiro Miki, Toshiya Sakoda, Yuta Hagiwara. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 20.01.2025. 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

Comparing Digital Versus Face-to-Face Delivery of Systemic Psychotherapy Interventions: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Pieter Erasmus[1*], MA; Moritz Borrmann[1,2*], MA; Jule Becker[1], BA; Lars Kuchinke[3], PhD; Gunther Meinlschmidt[1,4,5,6], PhD

[1]Department of Clinical Psychology and Cognitive Behavioral Therapy, International Psychoanalytic University Berlin, Berlin, Germany

[2]Day Clinic for Cognitive Neurology, University Hospital Leipzig, Leipzig, Germany

[3]Psychological Methods and Evaluation, International Psychoanalytic University Berlin, Berlin, Germany

[4]Department of Clinical Psychology and Psychotherapy – Methods and Approaches, Trier University, Trier, Germany

[5]Department of Digital and Blended Psychosomatics and Psychotherapy, Psychosomatic Medicine, University Hospital and University of Basel, Basel, Switzerland

[6]Department of Psychology, University of Basel, Basel, Switzerland

[*]these authors contributed equally

Corresponding Author:
Gunther Meinlschmidt, PhD
Department of Clinical Psychology and Cognitive Behavioral Therapy
International Psychoanalytic University Berlin
Stromstrasse 3b
Berlin, 10555
Germany
Phone: 49 30 300 117 710
Email: gunther.meinlschmidt@ipu-berlin.de

Abstract

Background: As digital mental health delivery becomes increasingly prominent, a solid evidence base regarding its efficacy is needed.

Objective: This study aims to synthesize evidence on the comparative efficacy of systemic psychotherapy interventions provided via digital versus face-to-face delivery modalities.

Methods: We followed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for searching PubMed, Embase, Cochrane CENTRAL, CINAHL, PsycINFO, and PSYNDEX and conducting a systematic review and meta-analysis. We included randomized controlled trials comparing mental, behavioral, and somatic outcomes of systemic psychotherapy interventions using self- and therapist-guided digital versus face-to-face delivery modalities. The risk of bias was assessed with the revised Cochrane Risk of Bias tool for randomized trials. Where appropriate, we calculated standardized mean differences and risk ratios. We calculated separate mean differences for nonaggregated analysis.

Results: We screened 3633 references and included 12 articles reporting on 4 trials (N=754). Participants were youths with poor diabetic control, traumatic brain injuries, increased risk behavior likelihood, and parents of youths with anorexia nervosa. A total of 56 outcomes were identified. Two trials provided digital intervention delivery via videoconferencing: one via an interactive graphic interface and one via a web-based program. In total, 23% (14/60) of risk of bias judgments were *high risk*, 42% (25/60) were *some concerns*, and 35% (21/60) were *low risk*. Due to heterogeneity in the data, meta-analysis was deemed inappropriate for 96% (54/56) of outcomes, which were interpreted qualitatively instead. Nonaggregated analyses of mean differences and CIs between delivery modalities yielded mixed results, with superiority of the digital delivery modality for 18% (10/56) of outcomes, superiority of the face-to-face delivery modality for 5% (3/56) of outcomes, equivalence between delivery modalities for 2% (1/56) of outcomes, and neither superiority of one modality nor equivalence between modalities for 75% (42/56) of outcomes. Consequently, for most outcome measures, no indication of superiority or equivalence regarding the relative efficacy of either delivery modality can be made at this stage. We further meta-analytically compared digital versus face-to-face delivery modalities for attrition (risk ratio 1.03, 95% CI 0.52-2.03; P=.93) and number of sessions attended (standardized mean difference

XSL•FO
RenderX

–0.11; 95% CI –1.13 to –0.91; *P*=.83), finding no significant differences between modalities, while CIs falling outside the range of the minimal important difference indicate that equivalence cannot be determined at this stage.

Conclusions: Evidence on digital and face-to-face modalities for systemic psychotherapy interventions is largely heterogeneous, limiting conclusions regarding the differential efficacy of digital and face-to-face delivery. Nonaggregated and meta-analytic analyses did not indicate the superiority of either delivery condition. More research is needed to conclude if digital and face-to-face delivery modalities are generally equivalent or if—and in which contexts—one modality is superior to another.

Trial Registration: PROSPERO CRD42022335013; https://tinyurl.com/nprder8h

(Interact J Med Res 2025;14:e46441) doi:10.2196/46441

KEYWORDS

systemic psychotherapy; family therapy; adolescent; systematic review; meta-analysis; face to face; digital; remote; distance; telehealth; delivery modality

Introduction

Background

Digital delivery of mental health interventions has gained increasing prominence in recent decades. Digital delivery provides solutions to many mental health intervention obstacles, such as limited access [1], geographical distance [2], financial constraints [3], and transportation issues [4]. These factors, and not least the COVID-19 pandemic [5], have contributed to increased delivery of digital modalities of mental health care [6]. One current line of research uses randomized controlled trials (RCTs) to determine the efficacy of digitally delivered family therapies [6] and parenting interventions [7], including but not limited to systemic approaches. This research supports the efficacy of the digital delivery of interventions and finds no evidence for digitally delivered interventions being inferior to face-to-face interventions [8-13]. However, most of the current literature does not compare face-to-face and digitally delivered interventions directly, does not differentiate between the types of digital delivery modalities (eg, therapist-guided videoconferencing and self-guided web-based programs), and does not consider how the modality differentially impacts specific disorders, outcomes, or populations. Furthermore, these studies aggregate findings across broad and diverse contextual definitions of family therapy and parenting interventions instead of applying unified and stringent theory-informed definitions. Although these reviews (eg, McLean et al [6] and Florean et al [7]) do not exclusively focus on a unified and stringent definition, they do provide evidence for systemic approaches in the broad context of family interventions. Many family therapists consider themselves to be systemic therapists, and there is an overlap between family therapy and systemic therapy (ST). However, ST is grounded in a theoretical treatment model, and unlike family therapy, it is not primarily defined by the therapy setting [14]. Thus, ST provides a unified and stringent definition that has practical consequences for research and clinical practice.

ST is a conceptual framework for mental health interventions that focuses on interpersonal relations, interactions, social surroundings and resources, perspectives, and constructions of problems; attempted solutions are appreciated and used as integral parts of the intervention [13,15,16]. Consequently, most family therapy and parenting interventions can be said to contain elements of ST to varying degrees while also potentially including elements from other conceptual frameworks, such as cognitive behavioral therapy and psychoeducation [17]. Conversely, as ST interventions often focus on families as important resources requiring incorporation into intervention delivery [18], they can, but need not be, similar to other family therapy or parenting interventions in terms of including other family members in their setting. The efficacy of ST as a face-to-face intervention is well documented in various reviews and meta-analyses for a range of psychological disorders for youth and adult populations [13,15-17,19]. In line with developments for other psychotherapeutic interventions, the implementation of digital delivery modalities of ST has rapidly increased in recent years [20-22]. The implementation of videoconferencing is particularly promising, and the experiences of practitioners implementing this delivery modality are overall positive [23]. Despite such indications, there are gaps in the literature [20,24]. There is an urgent need to provide an evidence base for practitioners to understand how face-to-face and digital delivery of the same ST intervention compare (compare McLean et al [6], Florean et al [7], and Fairburn and Patel [25]).

The need for greater granularity regarding the comparative efficacy of digital and face-to-face delivery modalities is in line with the more general requirement to identify “conditions under which systemic therapy works best” [16]. Thus, we simultaneously address both the need for a systematic review and meta-analysis that investigates (1) how various types of ST interventions, delivered in their traditional face-to-face form, compare to the same ST intervention delivered digitally with regard to their efficacy and (2) delivery modality as key contextual factors that modulate ST intervention outcomes.

Research Questions

We conducted a systematic review and meta-analysis to address the following research questions: (1) What are the characteristics, quality, and resulting evidence from published RCTs comparing the efficacy of ST interventions using digital treatment modalities with ST interventions using face-to-face treatment modalities across settings, populations, and outcomes? (2) If a meta-analysis of RCTs comparing ST interventions in digital and face-to-face modalities can be conducted, what is the difference in delivery modalities for mental disorder outcomes?

XSL•FO
RenderX

Methods

Overview

This systematic review followed reporting guidelines as laid out in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 statement (Multimedia Appendix 1 2020 checklist) [26]. We aimed to identify, evaluate, and synthesize findings from RCTs comparing digital and face-to-face delivery modalities for ST interventions across settings, target populations, and outcomes. This systematic review was preregistered (PROSPERO ID: CRD42022335013), with no deviation from the registered protocol.

Inclusion and Exclusion Criteria

We applied the following inclusion and exclusion criteria for articles to be considered for our systematic review within the following categories: (1) study design, (2) participants, (3) interventions, (4) digital delivery and face-to-face delivery modalities, (5) outcomes, and (6) article type.

RCTs of any type were included (eg, parallel, cluster, and quasi-RCTs). No minimum number of participants was prespecified.

Only those articles were deemed eligible for inclusion that involved individual participants of all ages and genders (1) with one or more diagnoses of a mental disorder, (2) identified as at risk in one or more mental or behavioral health domains, or (3) adversely affected by salient circumstances (eg, belonging to a structurally disadvantaged community and caring for a family member). Articles involving participants at the system level (eg, families) were also included as long as 1 member of the system fulfilled at least one of the aforementioned criteria. No further exclusion criteria were applied to increase the number of eligible studies.

We only included articles reporting on interventions delivered via at least 2 distinct delivery modalities: digital delivery and face-to-face delivery. Digital delivery was defined as exclusively relying on an external electronic medium (eg, videoconferencing hardware and software, telephone, and synchronous and asynchronous text-based conversations) for intervention delivery. While this allowed heterogeneity of digital delivery modalities across some dimensions (eg, synchronicity and extent of interaction with mental health practitioners), it kept other dimensions stable across all digital delivery modalities (use of electronic medium and no mental health practitioner physically copresent with recipients of intervention). Face-to-face delivery was defined as exclusively relying on in-person intervention delivery. This allowed for the use of external tools or media to supplement the intervention while specifying that intervention delivery had to occur solely between people physically present in the same physical space (eg, a clinician's office).

To increase the number of eligible studies, any relevant mental, behavioral, or somatic health efficacy; adherence and attrition; and further outcomes were included in this review. No a priori categorization of outcomes was applied. Efficacy outcomes were defined as any outcomes pertinent to improvement in relevant mental, behavioral, and somatic health domains. *Further outcomes* were defined as any outcomes not directly pertinent to improvement in relevant health domains (eg, subjective satisfaction rating of program delivery).

There were no restrictions imposed on the article type in terms of language or date. Articles published in peer-reviewed journals as full-text articles and other types of peer-reviewed full-text articles were included. Unpublished data (eg, dissertations), as well as study protocols, clinical trial registries, and non–peer-reviewed articles, were excluded.

Intervention Inclusion Criteria

Only articles reporting on ST interventions delivered via both digital and face-to-face delivery modalities were included. ST was operationalized based on 4 adapted definition criteria of ST [13], which were applied to the descriptions of the identified constituent primary parts (CPPs; eg, core sessions) of interventions. The definition criteria were derived from the following definition of ST (referred to as *systemic therapy*) with our adjustments in square brackets [13]:

> *We use systemic/systems-oriented therapy/therapies (ST) as a general term for a major therapeutic orientation that can be distinguished from other major approaches (e.g., CBT or psychodynamic therapy). We define systemic therapy as a form of psychotherapy that (1) perceives behavior and mental symptoms within the context of the social systems people live in; (2) focuses on interpersonal relations and interactions, social constructions of realities, and[/or] the recursive causality between symptoms and interactions; (3) includes family members and[/or] other important persons (e.g., teachers, friends, professional helpers) directly or indirectly through systemic questioning, hypothesizing, and specific interventions; and (4) appreciates and utilizes clients' perspectives on problems, resources, and[/or] preferred solutions.*

For each intervention, the appropriate unit of primary intervention delivery (as opposed to parts of the intervention labeled *supplementary*, *additional*, or *optional*, etc) was determined following the study's authors (eg, core sessions, core components, etc.). The total number of the intervention's CPPs was determined. If the intervention description did not afford to determine an appropriate unit of intervention delivery, the total number of CPPs was defined as 1.

For each description of content for any CPP, we assessed if it met any of our definition criteria of ST. Notably, any particular definition criterion could, in theory, be met by a set of CPP descriptions >1, allowing different content descriptions to meet the same criterion. This allowed a range of different interventions to be classified as being similar in kind (ie, being ST interventions), reflecting the integrative nature of clinical ST practice [27,28], rather than requiring all interventions to have identical CPPs.

All definition criteria of ST needed to be met by the respective CPPs, with at least 2 of the 4 definition criteria being met completely by all CPPs. If any of the definition criteria of ST were not completely met, at most 2 of the 4 definition criteria needed to be met by a minimum of at least 50% of all CPPs (or,

XSL•FO
RenderX

in the case of uneven numbers of CPPs, the closest possible number below the theoretical half point). The same CPP could satisfy >1 criterion. Definition criterion (1) could additionally be met by the relevant conceptual background provided.

Textbox 1 illustrates an example for 1 trial included in our study. Our approach used to screen trials against inclusion criteria for all trials is detailed in Multimedia Appendix 2 [13,29-32].

Textbox 1. Identified constituent primary parts (CPPs) of 1 trial against systemic therapy (ST) definition criteria for Behavioral Family Systems Therapy for Diabetes (BFST-D) intervention [33].

Number and unit of CPPs and 4 primary intervention components

1. Perceives behavior and mental symptoms within the context of the social systems people live in
   - Family functioning and maladaptive parent-child interactions were identified as central barriers to diabetes treatment adherence.
2. Focuses on interpersonal relations and interactions, social constructions of realities, or the recursive causality between symptoms and interactions: 3 of 4 CPPs
   - CPP1: family-problem solving (as a family, defining the problem, generating solutions, making decisions, implementing and monitoring results, and refining ineffective solutions)
   - CPP2: communication training (instruction, feedback, modeling, and rehearsal of approaches toward improving maladaptive communicative patterns)
   - CPP4: family restructuring (functional and structural approaches toward changing maladaptive or ineffective family system patterns and characteristics, such as weak parental coalitions or cross-generational coalitions).
3. Includes family members and other important persons (eg, teachers, friends, professional helpers) directly or indirectly through systemic questioning, hypothesizing, and specific interventions
   - All CPPs are delivered to the caregiver-adolescent dyad.
4. Appreciates and uses clients' perspectives on problems, resources, and preferred solutions: 2 of 4 CPPs are as follows:
   - CPP1 and CPP3: cognitive restructuring (addressing beliefs, attitudes, and attributions that could negatively affect effective interactions).

Search Strategy

Database Search

The following databases were searched from the earliest available date until March 15, 2022: PubMed, Embase (via Ovid), Cochrane CENTRAL (via Ovid), CINAHL (via EBSCO), PsycINFO (via EBSCO), and PSYNDEX (via EBSCO). The search strategy followed the guideline from Bramer et al [34], and the original search string was developed for use in PubMed and adapted manually for all other databases. The search strings contained a set of the following: (1) 1 compiled term for intervention type, (2) 1 compiled specificity term to define the target domain, and (3) 1 compiled delivery modality term. For the identification of RCT study design, we used the RobotSearch AI RCT highly sensitive filter [35], which provides high sensitivity in identifying RCTs [35-37]. Multimedia Appendix 3 [8,13,15,38-46] provides the full search strings, references of the published sources upon which we based our search terms, and explanations of the logical structure. On March 17, 2022, we conducted complementary searches in ClinicalTrials.gov, the International Clinical Trials Registry Platform, and the EU Clinical Trials Register (all via the respective standard website interface). On May 4, 2022, backward citation screening (articles cited) was conducted on identified relevant reviews and meta-analyses (Multimedia Appendix 4 [9,10,12,13,15,47-53]). On May 10, 2022, forward citation screening (articles citing included articles) was conducted via ISI Web of Science, backward citation screening on included articles was conducted via PubMed, and a "related articles" screening was conducted on included articles (first 20 articles listed as related articles) via PubMed.

Article Selection and Screening

Duplicates were removed using Automated Systematic Search Deduplicator [54]. The resulting list of search results was split into 2 sets (sorting by title in EndNote [version 20; Clarivate] and allocating alternating sets of approximately 100 consecutive references; number of articles=1816 and 1817) that were each assigned to 2 pairs of authors (PE and JB; MB and JB), with each author independently screening titles and abstracts using manualized standard operating procedures that we developed accommodating our inclusion and exclusion criteria. All discrepancies at the title and abstract screening stage were resolved through discussion between all 3 involved authors (PE, MB, and JB). PE and MB independently screened at the full-text stage. Discrepancies at this stage were resolved through discussion between PE and MB. GM was consulted as an independent author of this study in cases of persisting discrepancies after discussions at any stage. Because on several occasions a trial underlying an included article contributed data to >1 included article, article-trial correspondence was determined either via trial registration numbers (for 7 articles) or via references to articles on the same trial, unique trial names, as well as overlap in article authors, article dates, sample sizes and characteristics, interventions, and procedures (for 5 articles).

Data Extraction

Data were independently extracted by PE and MB using the Cochrane data collection form [55] for intervention reviews

XSL•FO
RenderX

(Multimedia Appendix 5 [29,30,33,56-59]). Discrepancies were resolved through discussion.

Risk of Bias

The authors (PE and MB) independently assessed the risk of bias (RoB) using the revised Cochrane Risk of Bias (version 2) tool for randomized trials [60] and following the process outlined in chapter 8 of the Cochrane handbook [55]. Discrepancies were resolved through discussion. RoB figures were created using *robvis* [61].

Data Synthesis and Analysis

We used RevMan (version 5.4) [62] to create forest plots and calculate, depending on the variable type and type of analysis, mean differences (MDs), standardized MDs, risk ratios, and associated CIs at the 95% level. Where possible, we calculated missing SDs using the Cochrane collaboration RevMan calculator [62]. We conducted meta-analyses only in cases of acceptable levels of clinical and methodological diversity and where sufficient data were available [63]. We conducted meta-analyses using random effects models as we expected heterogeneity across articles to be high. Statistical heterogeneity was assessed using the I^2 statistic. MDs and CIs were calculated where meta-analytic analysis was not possible, but nonaggregated analysis was still deemed appropriate. This was to increase qualitative comparability across studies with nonnegligible heterogeneity in statistical approaches and reported results. All relevant reported results are listed in Multimedia Appendix 6 [29,30,33,56-59,64-68]. We grouped MDs of outcomes according to their CIs into (1) fully located within the range from –0.20 to 0.20 (*equivalence*), (2) fully located outside this range (*superiority*), or (3) partially located within and partially located outside this range (*neither equivalence nor superiority*). In the absence of an established a priori minimal important difference, we accept Cohen *d* cutoff of 0.2 (small effect) as the defined range from –0.20 to 0.20 to indicate equivalence [69]. We interpreted a CI of an MD falling within, but not exceeding, this range as evidence for equivalence, a CI of an MD falling entirely outside this range as evidence for nonequivalence or superiority of one delivery modality, and a CI falling within and exceeding this range as insufficient evidence to determine equivalence or nonequivalence on the respective measure. Where calculating MDs and CIs was impossible due to missing information, reported results on differences between digital and face-to-face delivery modalities were provided (Multimedia Appendix 6). Study authors were contacted to provide missing information. A funnel plot was not created because we identified <10 trials [70].

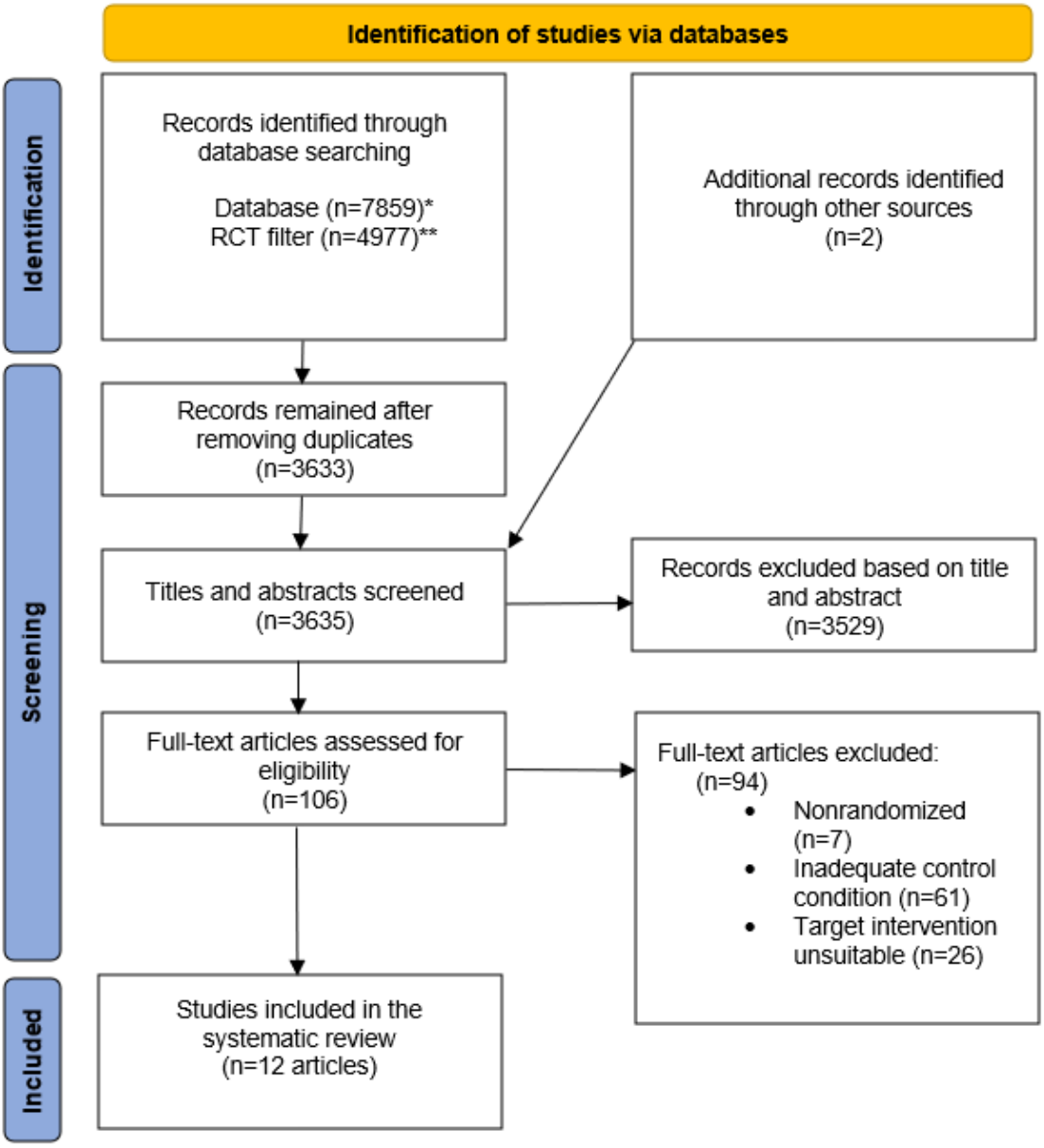
Results

Search Results

A flowchart of our search and screening process is depicted in Figure 1 [35]. The primary search for this study retrieved 4977 references and 3633 after removal of duplicates. Additional references (n=2) were identified through complimentary searches. The resulting 3635 references were screened at the title and abstract level. On the basis of our inclusion and exclusion criteria, 97.1% (3529/3635) articles were excluded during the title and abstract screening, and 2.9% (106/3635) articles were screened at the full-text screening stage. Overall, 0.3% (12/3633) of the initially identified articles reporting on 4 trials were included in this systematic review. Some (3/12, 25%) of the articles reporting on 2 trials [29,56,57] are only reported in Table 1 and Multimedia Appendix 6, as they could not be included in further meta-analytic aggregation or discrete analysis of MDs and CIs.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the articles included in the systematic review. RCT: randomized controlled trial. *Results from CINAHL (EBSCO): n=432; PSYNDEX (EBSCO): n=53; PsycINFO (EBSCO): n=505; Embase (Ovid): n=2177; Cochrane (Wiley): n=533; PubMed: n=4159; **RobotSearch AI, Marshall et al [35].



XSL•FO
RenderX

Table 1. Characteristics of included studies (N=754)[a].

Trial and articles	Trial sample size (% of N; F2F[b]; DD[c])	Youth age (y), mean (SD)	Male, n (%)	Feature	DI[d]	F2FI[e]	Sessions, mean (SD)
BFST-D[f]							
Duke et al [29], 2016	90 (11.9%; 46; 44)	15.02 (1.75)	55 (61.1)	HbA_{1c}[g] >9%	Videoconferencing	Clinic	5.8 (3.3)
Harris et al [58], 2015	90 (11.9%; 46; 44)	14.9[h] (1.7)	55 (61.1)[h]	HbA_{1c} >9%	Video conferencing	Clinic	5.8 (3.3)
Freeman et al [33], 2013	92 (12.2%; 45; 47)	15.1[i]	42 (45.7)[h]	HbA_{1c} >9%	Videoconferencing	Clinic	—[j]
Riley et al [57], 2015	82 (10.9%)[k]	14.1	55 (61.1)	HbA_{1c} >9%	Videoconferencing	Clinic	6.3 (3.4)
PAAS[l]							
Murry et al [56], 2019	421 (55.8%; 141; 141; CC[m]: 136)	—	195[h] (46)	PRB[n]	II[o]	GS[p]	—
Murry et al [59], 2019	412 (54.6%; 137; 138; CC: 137)	11.0	191[h] (46)	PRB	II	GS	—
Murry et al [30], 2018	412 (54.6%; 137; 138; CC: 137)	11.4	191 (46.4)	PRB	II	GS	4.0 (3.0)
F-PST[q]							
Kurowski et al [65], 2020	150 (19.9%; 34; 56; SG[r]: 60)	16.5 (1.1)	96 (64)	TBI[s]	Videoconferencing+OP[t,u]	Clinic	6.3 (2.6)
Wade et al [66], 2019	150 (19.9%; 34; 56; SG: 60)	16.5 (1.1)	96 (64)	TBI	Videoconferencing+OP[u]	Clinic	6.3 (2.6)
Wade et al [68],2019	150 (19.9%; 34; 56; SG: 60)	16.5 (1.1)	96 (64)	TBI	Videoconferencing+OP[u]	Clinic	6.3 (2.6)
Wade et al [67],2019	150 (19.9%; 34; 56; SG: 60)	15.5 (1.5)[v]	96 (64.0)	TBI	Videoconferencing+OP[u]	Clinic	6.3 (2.6)
SUCCEAT[w]							
Truttmann et al [64], 2020	102 (13.5%; 50; 52)	14.9 (1.9)	9 (8.9)	AN[x]	OP	GWS[y]	6.5 (2.0)

[a]Sum of n at the trial level. In case of inconsistencies in reported n across individual articles for each trial, the mode n was selected.

[b]F2F: face-to-face delivery.

[c]DD: therapist-guided digital delivery.

[d]DI: digital implementation.

[e]F2FI: face-to-face implementation.

[f]BFST-D: Behavioral Family Systems Therapy for Diabetes.

[g]HbA_{1C}: glycated hemoglobin.

[h]Calculated based on percentages and overall sample size provided.

[i]Calculated based on mean ages, sample sizes, and SDs provided.

[j]Not available.

[k]Only participants who completed the Child Depression Inventory measure (only those aged <18 y eligible).

[l]PAAS: Pathways for African American Success.

[m]CC: control condition.

[n]PRB: potential risk behavior.

[o]II: interactive interface.

[p]GS: group session.

[q]F-PST: Family Problem-Solving Therapy.

[r]SG: self-guided digital delivery.

XSL•FO
RenderX

[s]TBI: traumatic brain injury.

[t]OP: online program.

[u]Two distinct digital delivery conditions: (1) videoconferencing+web-based program and (2) online program.

[v]Most likely a clerical mistake in reporting.

[w]SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria.

[x]AN: anorexia nervosa.

[y]GWS: group workshop.

Characteristics of Included Articles

Number of Trials, Articles, and Sample Size

A total of 12 articles reporting on 4 trials were identified. One trial was reported on by a single article [64] (Supporting Carers of Children and Adolescents with Eating Disorders in Austria [SUCCEAT] trial), while the remaining 92% (11/12) of articles reported on a total of 3 trials. Of all 12 articles, 58% (n=7) of articles provided trial registration numbers, allowing for formally confirmed identification of the reported trial.

For the purposes of this study, the 4 trials are abbreviated according to their respective intervention names as follows: Behavioral Family Systems Therapy-Diabetes (BFST-D) trial [29,33,57,58], Family Problem-Solving Therapy (F-PST) trial [65-68], Pathways for African American Success (PAAS) trial [30,56,59], and the SUCCEAT trial [64].

As articles reporting on the BFST-D trial and articles reporting on the PAAS trial varied with regard to sample sizes (Table 1 and Multimedia Appendix 6), the total number of randomized participants across all trials could not be unambiguously determined. Given that for both trials, only 1 article each reported a different sample size as compared to all other articles reporting on the same trial, we reported the mode sample size for the total number of participants randomized to all conditions in all trials, including a control condition (N=754), and for the total number of participants randomized to face-to-face and digital delivery conditions (n=617). The same principle was applied for all reported participant characteristics, which varied in terms of proportions and absolute numbers depending on variation in reported sample sizes.

Participants

Some (4/12, 33%) articles reported on interventions delivered to youths with poorly controlled type 1 diabetes and their parents (BFST-D trial), 25% (n=3) articles reported on interventions delivered to parents, siblings, and youths identified as being at disproportionate risk for contracting HIV and sexually transmitted infections, teen parenthood, and substance use (PAAS trial). Moreover, 33% (4/12) articles reported on interventions delivered to youths with traumatic brain injury and their parents and other family members (F-PST trial), and 8% (1/12) articles reported on interventions delivered to parents of youths with anorexia nervosa (SUCCEAT trial). As there is inconsistent terminology between and within articles, the term *parent* is used here and throughout this paper to refer to parents, grandparents, legal guardians, and caregivers included in the respective articles. All identified interventions either involved families in intervention delivery (BFST-D, PAAS, and F-PST trials) or in terms of interpersonal target outcomes (SUCCEAT trial). Additional participant information is provided in Table 1 and Multimedia Appendix 7 [29,30,33,56-59,64-68]. Participant characteristics were reported to be similar across conditions in all articles, with the exception of the F-PST trial, which included a lower proportion of White participants in the face-to-face delivery condition, probably linked to quasi-randomization involving place of residence in the allocation procedure.

Face-to-Face and Digital Interventions and Delivery Format

Interventions were provided by health professionals in 75% (9/12) of articles (BFST-D, F-PST, and SUCCEAT trials) and by trained community leaders in 25% (3/12) articles (PAAS trial). Intervention content was similar across delivery modalities for all trials, with all articles explicitly mentioning the same intervention being delivered across conditions and all CPPs of interventions being delivered in both conditions (Table 1 and Multimedia Appendix 2).

All therapist-guided digital delivery was provided via videoconferencing software. Self-guided intervention delivery was provided via access to web-based modules for 33% (4/12) of the articles (F-PST trial) and for 25% (3/12) of the articles (PAAS trial) via an interactive interface visually representing users through avatars in a virtual environment. Face-to-face delivery was reported as individual in-clinic sessions for 67% (8/12) of the articles (BFST-D and F-PST trials), as group sessions at community centers for 25% (3/12) of the articles (PAAS trial), and in-clinic workshops for multiple individuals for 8% (1/12) of the articles (SUCCEAT trial).

Outcomes

Across all 12 articles, 56 outcome measures (ie, dependent variables) were reported, including cases of parent-report and youth-report measures of the same construct. Primary outcomes (8/56, 14%), 71% (40/56) secondary outcomes, and 14% (8/56) further, dropout, or attrition outcomes were identified. We have provided an overview of all identified outcomes in Multimedia Appendix 8 [29,33,56-59,64-68].

We summarized and structured data along three domains: (1) outcome type (ie, youth, parent, family functioning, and dropout, attrition, or further), (2) delivery modality type (face-to-face, therapist-guided digital, and self-guided digital), and (3) measurement time point (postintervention test and follow up test) to reduce clinical and methodological diversity within outcome data categories. We provide a more detailed overview of delivery modality types (eg, videoconferencing or in-clinic sessions) in Table 1, an overview of outcome measures (ie, dependent variables) in Multimedia Appendix 8, and measurement time points in Multimedia Appendix 6.

XSL•FO
RenderX

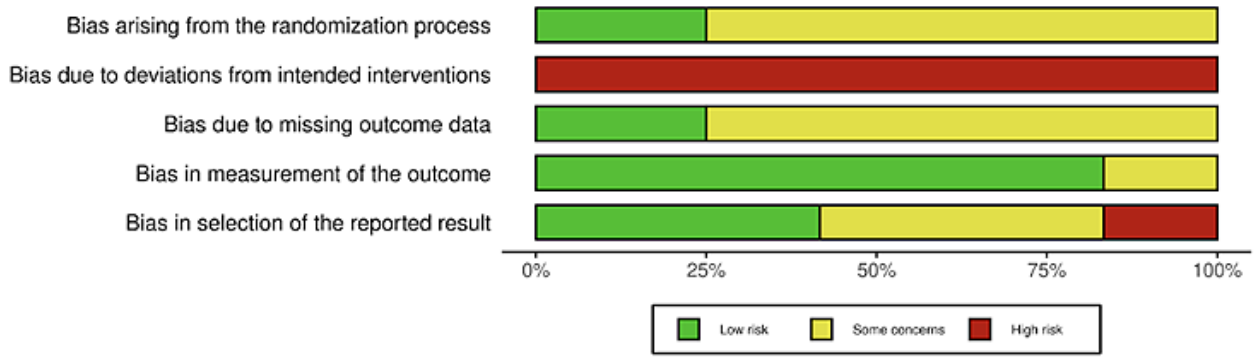
The application of these 3 domains to the categorization of data yielded 16 potential outcome categories: youth (Multimedia Appendix 9 [58,59,64-67]); parent (Multimedia Appendix 9); family functioning (Multimedia Appendix 10 [59]); and adherence, attrition, and further outcomes (Multimedia Appendix 11 [30,33,62,65,68]) for either therapist-guided digital versus face-to-face delivery (Multimedia Appendix 9) or self-guided digital versus face-to-face delivery modalities (Multimedia Appendices 9 and 10) at either posttest (Multimedia Appendices 9 and 10) or follow-up test (Multimedia Appendix 11) time points.

Risk of Bias

Overview

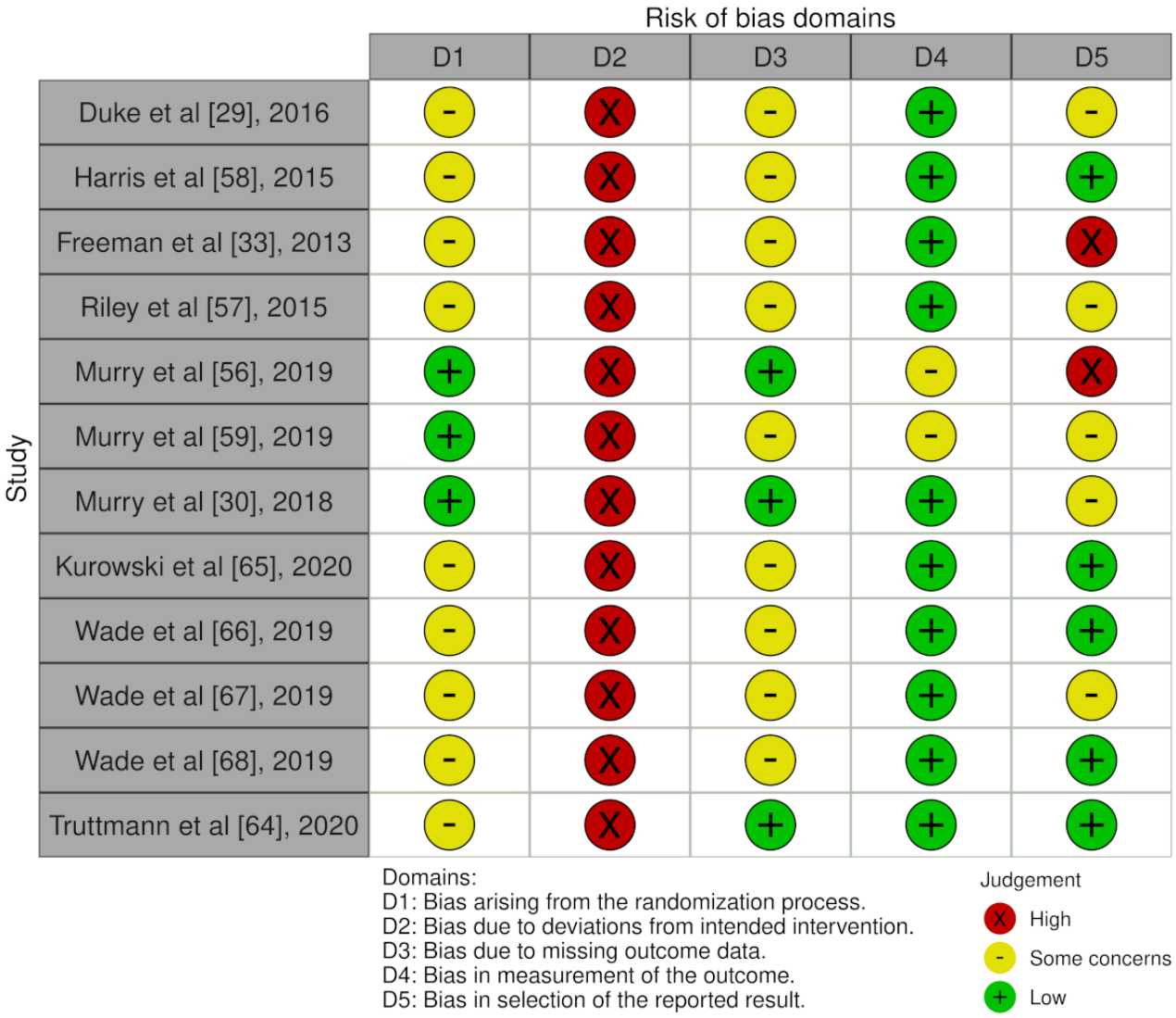
The 12 articles were rated as high risk for allocation concealment. For sequence generation, 75% (9/12) articles were rated as having some concerns. For randomization, 2 of the 4 trials used quasi-randomization. Some (9/12, 75%) articles were rated as having some concerns in terms of missing data. Overall proportions of RoB levels across articles are presented in Figure 2; individual RoB assessments at the article level are presented in Figure 3 [29,30,33,56-59,64-68]. Moreover, 92% (11/12) articles reported on 3 trials, with inconsistent reporting at times, which constituted a challenge for assessing the RoB, particularly for the domain of selective outcome reporting, as multiple articles reporting on the same trial were found to simultaneously overlap in some reported outcomes and diverge for other outcomes (eg, inconsistent reporting of the Diabetes Family Conflict Scale across articles reporting on the BFST-D trial).

Figure 2. Total proportions of articles' risk of bias assessments by assessment domain.



XSL•FO
RenderX

Figure 3. Risk of bias assessments of all included articles. D1: bias arising from the randomization process; D2: bias due to deviations from intended intervention; D3: bias due to missing outcome data; D4: bias in measurement of the outcome; D5: bias in selection of the reported result.



Sequence Generation

Of all 12 articles, 8% (n=1) reported a quasi-randomization temporal sequencing design (SUCCEAT trial); 33% (n=4) reported a quasi-randomized, multicenter RCT design (F-PST trial), in which participants were unequally allocated to the digital and face-to-face delivery conditions based on their place of residence (all participants who lived outside a 40-kilometer radius from the site were randomly allocated to the digital delivery conditions, and participants within the radius were randomly allocated to all conditions, favoring the face-to-face delivery condition by a ratio of 2:1); and 58% (n=7) reported randomized controlled trials (BFST-D and PAAS trials). Of those 7 articles, 57% (n=4) reported block randomization without providing information on allocation sequence concealment, increasing the RoB from sequence generation (BFST-D trial). Overall, 25% (3/12) of the included articles were rated low RoB arising from the randomization, while 75% (9/12) were rated as giving rise to some concerns.

Allocation Concealment

The question of adequate allocation blinding procedures (for participants and intervention providers) is a common concern in psychotherapy outcome research [71]. This also applies to the articles included in this review. Given the absence of a universally agreed-upon alternative approach [71], all articles were rated as high RoB in this domain.

Handling of Incomplete Data

For any of the 12 articles, not all outcome data were available. Some (2/12, 17%) of the articles provided evidence of data missing at random [56] or treated attrition as an outcome variable [30] and were rated as having low RoB. Most (9/12, 75%) articles did not adequately analyze if missing outcome data did not potentially bias the results and were rated as giving rise to some concerns.

XSL•FO
RenderX

Outcome Assessment

RoB arising from the measurement of the outcome variables was rated *low risk* for 83% (10/12) of the articles and with *some concerns* for 17% (2/12) of the articles that used a large proportion of outcome measures for which judgments regarding validity and reliability could not be made [56,59] (Multimedia Appendix 8).

Selective Reporting

As indicated earlier, RoB from selective reporting was challenging to assess given the interaction of multiple articles reporting on the same trial and inconsistencies across articles on the same trials. This was further exacerbated in the case of a trial for which no preregistration could be identified (PAAS trial) and of an article exclusively reporting on unregistered outcomes of a preregistered trial [33]. Some (5/12, 42%) articles were rated as giving rise to some concerns, 9% (5/12) were rated as having low RoB, and 17% (2/12) of the articles were rated as having high RoB.

Appropriateness of a Meta-Analysis to Determine the Efficacy of Delivery Modalities, Adherence, and Attrition

Meta-analytic aggregation was deemed inappropriate for all delivery modality comparisons of youth, parent, and family functioning outcomes due to the relatively low number of trials within the respective outcome categories, combined with the presence of substantial clinical and methodological diversity across trials [63]. Meta-analytic aggregation of the number of sessions attended and attrition was conducted because heterogeneity was lower and the number of eligible trials was higher for these outcomes as compared to the other outcome measures reported and included (Figure 4 [58,64,65] and Figure 5 [58,64,65]). The substantial clinical and methodological diversity across trials remains problematic for drawing meaningful conclusions using meta-analytic aggregations. However, a preliminary quantitative synthesis was included as it still serves as a starting point for discussion and future research.

Figure 4. Meta-analysis of number of sessions attended in face-to-face versus digital delivery modality comparison.

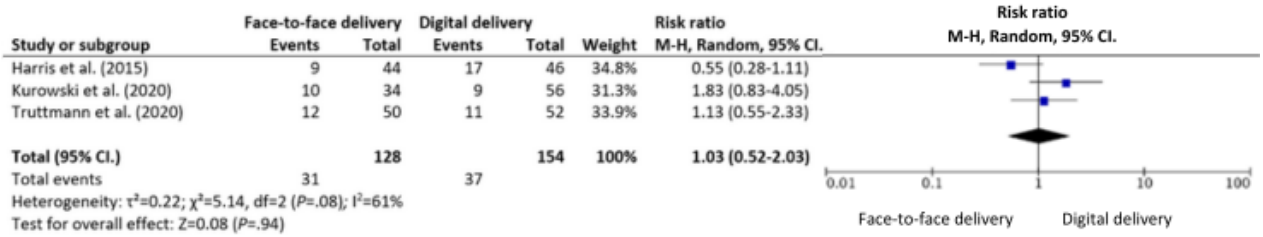
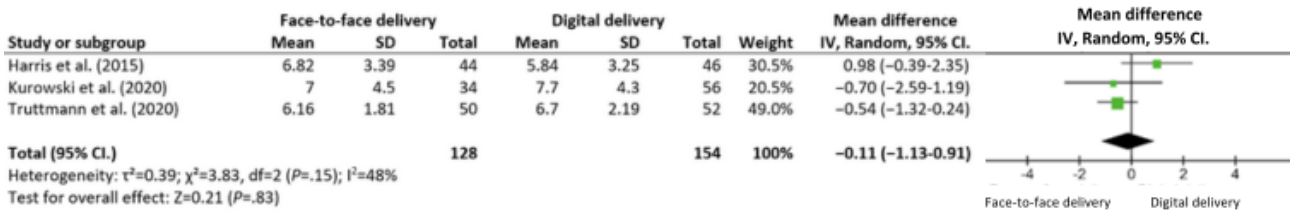


Figure 5. Meta-analysis of drop-out (event) risk ratio in face-to-face versus digital delivery modality comparison.



Nonaggregated Analyses for Face-to-Face and Digital Delivery Modalities

Of all 56 reported outcomes (ie, dependent variables) across all trials and articles, 25% (n=14) outcomes from 2 trials (13/14, 93% in the F-PST trial and 1/14, 7% in the SUCCEAT trial) showed MDs with corresponding CIs falling either (1) fully within our predefined range of minimal important difference of –0.20 to 0.20 or (2) fully outside of this range. These outcomes are reported in Table 2 and Multimedia Appendix 12. All remaining 75% (42/56) outcomes showed MDs with corresponding CIs partially located within and partially located outside this range. Among these, 4% (2/56) further outcomes, 2% (1/56) family functioning outcome, and 2% (1/56) youth outcome were reported as substantially different between delivery modalities in the original articles while still including our range for the minimal important difference. Multimedia Appendix 9 shows a full list of all MDs and CIs of outcomes.

XSL•FO
RenderX

Table 2. Outcomes with substantial differences (95% CI excluding 0) or equivalence across all trials (N=252; sum of n at the trial level).

Outcome type, postintervention test and follow-up test, trial, and article	Outcome	Type of DD[a]	F2F[b] delivery, n	F2F delivery, mean (SD)	DD, n	DD, mean (SD)	Mean difference (95% CI)	Directionality
Youth								
Postintervention test								
F-PST[c]; Wade et al [67], 2019								
	PedsQL[d], parent report	TG[e]	22	62 (15.71)	43	70.9 (18.49)	–8.90 (–17.48 to –0.32)	Favors DD
	PedsQL, parent report	SG[f]	22	62 (15.71)	51	71.7 (18.85)	–9.70 (–18.06 to –1.34)	Favors DD
Follow-up test								
F-PST; Wade et al [67], 2019								
	HBI[g] somatic, youth report	TG	22	9.8 (4.92)	44	6.6 (5.90)	3.20 (0.50 to 5.90)	Favors DD
	HBI cognitive, parent report	SG	22	19.1 (8.58)	48	12.3 (9.91)	6.80 (2.25 to 11.35)	Favors DD
	HBI somatic, youth report	SG	22	9.8 (4.92)	48	6.8 (5.89)	3.00 (0.35 to 5.65)	Favors DD
	PedsQL, parent report	SG	22	67 (16.18)	48	76.6 (18.98)	–9.60 (–18.23 to –0.97)	Favors DD
Parent								
Postintervention test								
F-PST; Wade et al [66], 2019								
	CES-D[h]	TG	22	15.1 (1.81)	43	12.8 (1.53)	2.30 (1.42 to 3.18)	Favors DD
	BSI[i]	TG	22	56.1 (2.46)	43	53.3 (2.02)	2.80 (1.61 to 3.99)	Favors DD
	CES-D	SG	22	15.1 (1.81)	51	13.5 (1.42)	1.60 (0.75 to 2.45)	Favors DD
	BSI	SG	22	56.1 (2.46)	51	54.7 (1.88)	1.40 (0.25 to 2.55)	Favors DD
SUCCEAT[j]; Truttmann et al [64], 2020								
	SCL[k]-90-R	TG	48	0.24 (0.26)	46	0.31 (0.36)	–0.07 (–0.20 to 0.06)	Equivalence
Follow-up test								
F-PST; Wade et al [66], 2019								
	BSI	TG	22	51.7 (2.42)	44	53.3 (2.02)	–1.60 (–2.77 to –0.43)	Favors F2F delivery
	CES-D	SG	22	11.8 (2.03)	48	13.2 (1.58)	−1.40 (−2.36 to −0.44)	Favors F2F delivery
Wade et al [68], 2019								
	P-PE[l]	TG	22	9.0 (1.4)	43	7.8 (2.1)	1.20 (0.34 to 2.06)	Favors F2F delivery

[a]DD: digital delivery.

[b]F2F: face-to-face.

[c]F-PST: Family Problem-Solving Therapy.

[d]PedsQL: Pediatric Quality of Life Inventory.

[e]TG: therapist guided.

[f]SG: self-guided.

XSL•FO
RenderX

[g]HBI: Health and Behaviour Inventory.

[h]CES-D: Center for Epidemiological Studies Depression Scale.

[i]BSI: Brief Symptom Inventory.

[j]SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria.

[k]SCL: Symptom Checklist 90-Revised Global Severity Index.

[l]P-PE: Parent Rated Program Evaluation.

Of the 14 outcomes falling fully within or fully outside the minimal important difference range, superiority of the digital delivery condition was found for 71% (n=10) outcomes. Of these 10 outcomes, 60% (n=6) were youth outcomes and 40% (n=4) were parent outcomes. The quality of life of youths, reported by parents (Pediatric Quality of Life Inventory, parent report) at post intervention test, was substantially higher for both therapist-guided and self-guided digital delivery modalities compared to face-to-face delivery modalities and remained substantially higher for the self-guided digital delivery condition at follow-up. Somatic symptoms of youths, reported by them (Health and Behaviour Inventory somatic, youth report) at follow-up, were substantially higher in the face-to-face delivery condition compared to both therapist-guided and self-guided digital delivery conditions. In addition, cognitive symptoms of youths, reported by parents (Health and Behaviour Inventory cognitive, parent report) at follow-up, were substantially higher in the face-to-face delivery condition compared to the self-guided digital delivery condition. Depressive symptoms of parents (Center for Epidemiological Studies Depression Scale) at post intervention test were substantially higher in the face-to-face delivery condition compared to both therapist-guided and self-guided digital delivery conditions. Similarly, mental symptoms of parents (Brief Symptom Inventor) at post intervention measurement were substantially higher in the face-to-face delivery condition compared to both therapist-guided and self-guided digital delivery conditions.

Superiority of the face-to-face delivery condition was found for 3 outcomes. All of these outcomes were parent outcomes from the F-PST trial at follow-up. In contrast to comparisons at post intervention measurement, depressive symptoms of parents (Center for Epidemiological Studies Depression Scale) at follow-up were substantially higher in the self-guided digital delivery condition compared to the face-to-face delivery condition. Similarly, contrasting with comparisons at post intervention measurement, mental symptoms of parents (Brief Symptom Inventory) at follow-up were substantially higher in the therapist-guided digital delivery condition compared to the face-to-face delivery condition. Program evaluation by parents was more favorable in the face-to-face delivery condition compared to the therapist-guided digital delivery condition.

Only 2% (1/56) of outcomes fell fully within the minimal important difference range. In the SUCCEAT trial, psychopathological symptoms of parents (Symptom Checklist 90-Revised Global Severity Index) at post intervention test were equal between therapist-guided digital and face-to-face delivery conditions.

Meta-Analytic Aggregation of Adherence and Attrition

A total of 9 articles reporting on all 4 trials (BFST-D trial [29,57,58], PAAS trial [59], F-PST trial, and SUCCEAT trial) reported on the number of sessions attended, and all (11/12, 92%) articles, except for 1 (PAAS trial [57]), reporting on all trials reported on attrition. The number of sessions attended and attrition were meta-analyzed, and the meta-analytic aggregations are reported in Figures 4 and 5. The heterogeneity between trials was moderate (I^2=48% for number of sessions attended) to high (I^2=61% for attrition), and given the low number of included trials, this heterogeneity estimate might be inaccurate [72], and conclusions from the meta-analyses should thus be interpreted with caution.

Discussion

Principal Findings

We identified 12 articles reporting on 4 trials from which data were extracted (Multimedia Appendix 5). Of note, despite broad inclusion criteria for participants, participants included only youths or parents of youths. Recipients of interventions were youths with poor diabetic control, traumatic brain injuries, increased risk behavior likelihood, and parents of youths with anorexia nervosa. A total of 56 relevant outcomes were identified. Two trials provided digital intervention delivery via videoconferencing, one via an interactive graphic interface, and the other via a web-based program. RoB judgments varied substantially across domains and articles, with largely *some concerns* judgments for randomization, missing outcome data, and selective reporting. Applying a minimal important difference criterion to indicate equivalence or superiority, nonaggregated analyses of MDs and CIs between delivery modalities yielded varying types of comparative outcomes across the included studies: superiority of the digital delivery modality and superiority of the face-to-face delivery modality (CIs falling fully outside the range of the minimal important difference) and equivalence between delivery modalities (CIs falling fully inside the range of the minimal important difference), with most comparative outcomes, indicating neither superiority of one modality nor equivalence between delivery modalities (CIs falling within and outside the range of the minimal important difference). Overall, our analyses did not reveal any strong evidence for the superiority of face-to-face compared to digital delivery conditions and might tentatively instead indicate a higher proportion of favorable effects of digital delivery modalities for certain outcomes and certain contexts. Due to the qualitative nature of this analysis, the limited number of studies included in this review, and the large heterogeneity between studies, results need to be interpreted with caution. In addition, more research is needed to draw meaningful conclusions on the comparative efficacy of digital versus face-to-face delivery modalities of ST interventions.

We found a limited and insufficient amount of evidence to meta-analytically estimate the comparative efficacy of digital

XSL•FO
RenderX

versus face-to-face delivery modalities of ST interventions with regard to most psychological, somatic, and behavioral parent and youth outcomes, as well as family functioning and further outcomes at either postintervention test or follow-up test. Meta-analytic analysis was deemed appropriate only for attrition and the number of sessions attended. We compared digital versus face-to-face delivery modalities for attrition and the number of sessions attended, showing no evidence for the superiority of either delivery modality in these domains.

Comparison With Prior Work

Our findings are largely consistent with related findings from a recent meta-analysis on family therapy, including 3 trials reporting no evidence for a difference at postintervention test between therapist-guided digital versus face-to-face delivery of interventions for youth outcomes and parent outcomes [6]. Similarly, our findings are consistent with findings from a recent meta-analysis on therapist-guided digital parenting interventions that included 4 trials reporting no evidence for a difference at posttest for youth outcomes between therapist-guided digital and face-to-face delivery conditions [7]. None of this prior work, including ours, either strongly supports the superiority of face-to-face over digital delivery conditions or vice versa.

This study adds to research as the first to contribute to the existing body of research from an ST standpoint comparing RCTs of digital and face-to-face delivery modalities of the same intervention. While showing considerable overlap with family therapy, couple therapy, and parenting interventions, ST interventions are characterized by a largely independent theoretical framework and set of therapeutic techniques. Recent reviews in related areas of research [6,7,10,47,73] include interventions with different conceptual backgrounds (eg, family-based cognitive behavioral therapy and the integrated family intervention for child conduct problems [6]) under umbrella terms, such as family therapy or parenting interventions. The theory-driven approach we adopted for this systematic review helps to increase granularity by only including interventions based on ST content and conceptual background.

Strengths and Limitations

Our systematic review has several strengths. First, we applied rigorous methodology, which was preregistered and followed PRISMA guidelines [26]. Second, we applied a search strategy designed for high sensitivity to 6 databases, with an extensive range of complimentary searches coupled with consistent application of inclusion criteria on study design, intervention type, and delivery modalities, which enabled the descriptive synthesis of published evidence based on diverse target populations, different contexts, and different digital delivery modalities. Third, including ST interventions by identifying CPPs and testing them against our definition criteria ensured high levels of reproducibility and validity.

However, there are also limitations to be considered when interpreting the findings of this study. First, the definition of ST applied here is an adapted version of one of several, at times inconsistent, definitions of ST currently used [74,75]. However, our definition is based on the widely accepted definition by von Sydow et al [19], which has also been used as the basis for large-scale public policy and insurance regulations. Second, the operationalized definition criteria were only applied to identified CPPs of the interventions and for criterion 1 (namely, "perceives behavior and mental symptoms within the context of the social systems people live in"), conceptual background to make inclusion or exclusion judgments during screening. This approach may have excluded additional potentially relevant information about the interventions not provided as part of the description of the intervention itself (eg, the information provided in cited articles). However, we judged this approach to be the most beneficial to this study's replicability by reducing the arbitrary selection of intervention features as much as possible. Third, given that this study was, to our knowledge, the first to investigate digital versus face-to-face delivery modalities for ST interventions, no a priori categorical constraints were defined with regard to participants, outcomes, or contexts. Instead, constraints were defined post hoc, with the goal of maximizing diversity between and minimizing diversity within subgroups. Of note, despite a few constraints regarding the type of participant population, the sample from which participants were recruited was found to be comparatively narrowly defined (youths or parents of youths). Fourth, while our definition of the minimal important difference is in line with current guidelines [19] and current research in related fields [69], also based on Cohen *d* cutoff of 0.2 for a small effect, there are other approaches, such as anchor-based estimates [76], which might yield different ranges for the minimal important differences for the scales included in this study. However, given the fact that the required data for other approaches toward defining the minimally important differences, such as anchors or expert ratings, are currently not available for all identified outcome measures, our approach based on a Cohen *d* cut-off is acceptable and feasible.

Generalizability

Some factors limit the generalizability of our findings. First, while the identification of a low number of trials constitutes a finding in its own right, it limits the appropriateness of meta-analytic aggregation. This was exacerbated by the fact that multiple articles with potentially relevant findings reporting on the same trial were identified and that some articles, in some cases, showed inconsistencies in reporting data (eg, on sample size). Second, despite post hoc introduction of categories, diversity within these categories remained substantial (eg, in terms of participants' race, differences in face-to-face delivery formats, or different outcome domains), further limiting the appropriateness of meta-analytic aggregation. The differential impact of the type of digital modality used in each trial further limits the results of our study, and due to the low number of trials, it was not possible to analyze these tools separately. Third, the high diversity between trials (eg, in terms of delivery modalities, patient features, outcomes, and contexts) combined with the low overall number of trials limits the generalizability of this study's findings to other populations and delivery contexts. Fourth, the low number of identified trials did not allow for subgroup analyses based on RoB judgments. Fifth, 3 of the 4 trials were conducted in the United States, and 3 of the 4 trials recruited a substantial majority of White participants, further limiting generalizability to other contexts and target

XSL•FO
RenderX

populations. Sixth, the influence of factors such as paraverbal communication, subject of discussion, and client-practitioner dynamics is crucial for understanding psychotherapeutic processes. Although RCTs are seen as the gold standard in evidence-based medicine, they may not adequately represent these factors, and the same applies to our study by extension.

Clinical and Research Implications

This review has some important clinical and research implications. Before digital interventions can be accepted unanimously and unambiguously, solid evidence is needed to be built on specific criteria for evaluating the appropriateness of digital modes of delivering ST interventions. Digital delivery is rapidly increasing despite insufficient evidence on the comparative efficacy of digital and face-to-face delivery modalities of ST. This might explain some of the reported clinicians' resistance to implementing digital interventions in their practices [77] and should caution against replacing face-to-face with digital delivery modalities without careful consideration. At the same time, given the evidence base supporting the efficacy of digitally delivered systemic psychotherapy (though currently still not as broad as that for face-to-face delivery), the largely inconclusive evidence regarding the comparative efficacy of delivery modalities paired with some indicators for the superiority of digital delivery modalities of ST for certain populations and contexts could also be interpreted as a caution against adopting face-to-face delivery of ST interventions as the clinical default. Our study might serve as a first tentative indicator that different populations benefit differently from different delivery modalities. One could, for instance, look at differences between significant outcomes from the included trials in our study and speculate if parents might be more likely to benefit from face-to-face delivery than youths who, in turn, might be more likely to benefit from digital delivery modalities. Potential avenues for future research in this regard could, for instance, investigate the impact of the parents' role as facilitators of self-guided group interventions. At this point, these points of departure for future research remain speculative, given the current lack of clear evidence regarding the comparative efficacy of digital and face-to-face delivery modalities of ST. To that end, more studies with equivalence design are required.

According to Greene et al [78], equivalence studies should include 4 key aspects in their design: an active comparator condition, identifying appropriate equivalence margins, calculating adequate sample size estimations, and doing appropriate statistical analysis [78]. Our review highlights the need for research implementing all these key aspects. Furthermore, future research should explore subgroups or conduct sensitivity analyses to understand the limitations and the differential advantages of digital and face-to-face delivery modalities. The questions, such as how different types of digital interventions compare to face-to-face interventions and which intervention is efficacious for target populations, specific contexts, and various modes of digital interventions, including therapist-guided, self-guided, videoconferencing, virtual augmented reality, and smartphone apps, remain unanswered. Moreover, research on digital apps and interventions that, even if not directly involved in therapy, may provide several benefits in supporting those delivering and receiving mental health interventions is required. There is a need to investigate digital modes of delivery that could potentially complement face-to-face interventions through blended forms of care (refer to the study by Erbe et al [79]), specifically in the context of ST. Overall, more high-quality controlled trials using similar outcome measures as well as systematic reviews and meta-analyses are required to scrutinize under what circumstances digital ST interventions work best, especially given the increasing amount of available modes of delivery and the increase in digital delivery of mental health care in general.

Conclusions

Taken together, our systematic review indicated that the current evidence regarding the comparative efficacy of digital and face-to-face delivery modalities of ST interventions does not allow conclusions to be drawn regarding equivalence or superiority for youth, parent, family functioning, adherence, attrition, or further outcomes. There is a dearth of RCTs comparing the efficacy of comparable ST interventions delivered via both digital and face-to-face delivery modalities. The low number of trials combined with high clinical and methodological diversity between trials did not permit drawing meaningful conclusions from meta-analytic aggregation, and neither equivalence nor superiority of one modality can be excluded at this point. Nonaggregated analyses of MDs and CIs between delivery modalities indicated that neither equivalence nor superiority of either modality can be excluded at this stage for most outcomes, target populations, and contexts. As digital delivery of ST interventions is becoming increasingly ubiquitous, more research is urgently required to investigate how digital delivery of ST interventions compares to face-to-face delivery of the same interventions. There is a need for more research with a particular focus on different types of digital delivery modalities, including blended forms of mental health care, investigating potential variation in efficacy across different settings, populations, outcomes, and contexts.

Acknowledgments

GM received funding from the Stanley Thomas Johnson Stiftung and Gottfried und Julia Bangerter-Rhyner-Stiftung (PC 28/17 and PC 05/18), from Gesundheitsförderung Schweiz (18.191/K50001), from the Swiss Heart Foundation (FF21101), from the Research Foundation of the International Psychoanalytic University Berlin (5087), from the Swiss National Science Foundation (100014_135328), from the Hasler Foundation (23004), in the context of a Horizon Europe project from the Swiss State Secretariat for Education, Research and Innovation contract 22.00094, and from Wings Health in the context of a proof-of-concept study. GM and LK received funding from the German Federal Ministry of Education and Research (budget item 68606). GM and MB

XSL•FO
RenderX

received funding from the Research Foundation of the International Psychoanalytic University Berlin (5217). The article was funded and supported by the Open Access Fund of Universität Trier and by the German Research Foundation (DFG).

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request. A list of articles excluded with respective justifications for exclusions at the full-text screening stage can be made available upon request.

Authors' Contributions

PE and MB conceptualized the study, conducted data analysis and interpretation, and wrote the manuscript draft. PE, MB, and JB contributed to data collection. PE, MB, GM, LK, and JB revised the manuscript. PE, MB, GM, LK, and JB contributed to the final approval of the manuscript for publication.

Conflicts of Interest

GM is the cofounder and holds stock in Therayou AG, which is active in the field of digital and blended mental health care. GM receives royalties from publishing companies as an author, including a book published by Springer, and an honorarium from Lundbeck for speaking at a symposium. Furthermore, GM is compensated for providing psychotherapy to patients, acting as a supervisor, serving as a self-experience facilitator (“Selbsterfahrungsleiter”), and for postgraduate training of psychotherapists, psychosomatic specialists, and supervisors. Other authors do not declare any conflicts.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[DOCX File , 34 KB - ijmr_v14i1e46441_app1.docx]

Multimedia Appendix 2
Intervention inclusion criterion and constituent primary parts of interventions meeting definition criteria of systemic psychotherapy.
[DOCX File , 25 KB - ijmr_v14i1e46441_app2.docx]

Multimedia Appendix 3
Search terms, structure of search string, and database adaptations.
[DOCX File , 34 KB - ijmr_v14i1e46441_app3.docx]

Multimedia Appendix 4
Data extracted.
[DOCX File , 19 KB - ijmr_v14i1e46441_app4.docx]

Multimedia Appendix 5
Reported statistics on face-to-face versus digital delivery.
[DOCX File , 39 KB - ijmr_v14i1e46441_app5.docx]

Multimedia Appendix 6
Characteristics of included articles.
[DOCX File , 52 KB - ijmr_v14i1e46441_app6.docx]

Multimedia Appendix 7
Outcome measures overview.
[DOCX File , 23 KB - ijmr_v14i1e46441_app7.docx]

Multimedia Appendix 8
Comparison tables for outcomes.
[DOCX File , 44 KB - ijmr_v14i1e46441_app8.docx]

Multimedia Appendix 9
Family functioning outcomes.
[DOCX File , 54 KB - ijmr_v14i1e46441_app9.docx]

XSL•FO
RenderX

Multimedia Appendix 10
Adherence, attrition, and further outcomes.
[DOCX File , 20 KB - ijmr_v14i1e46441_app10.docx]

Multimedia Appendix 11
Screened reviews and meta-analyses.
[DOCX File , 33 KB - ijmr_v14i1e46441_app11.docx]

Multimedia Appendix 12
Flowchart of participant characteristics, measurement time points, intervention delivery type, and superiority or equivalence of delivery modalities. BSI: Brief Symptom Inventory; CES-D: Center for Epidemiological Studies Depression Scale; DD: digital delivery; F2F: face-to-face; F-PST: Family Problem-Solving Therapy; HBI: Health and Behaviour Inventory; P-PE: Parent Rated Program Evaluation; SCL: Symptom Checklist 90-Revised Global Severity Index; SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria; PedsQL: Pediatric Quality of Life Inventory.
[PNG File , 1130 KB - ijmr_v14i1e46441_app12.png]

References

1. Dack C, Pal K, Ross J, Michie S, May C, Stevenson F, et al. Digital health interventions for adults with type 2 diabetes: qualitative study of patient perspectives on diabetes self-management education and support. J Med Internet Res 2018 Jan 29;20(2):e40 [FREE Full text] [doi: 10.2196/jmir.8439] [Medline: 29463488]
2. Pompili S, Orsolini L, Salvi V, Volpe U. A systematic review on TeleMental health in youth mental health: focus on anxiety, depression and obsessive-compulsive disorder. Medicina (Kaunas) 2021 Jul 31;57(8):793 [FREE Full text] [doi: 10.3390/medicina57080793] [Medline: 34440999]
3. Doss BD, Knopp K, Roddy MK, Rothman K, Hatch SG, Rhoades GK. Online programs improve relationship functioning for distressed low-income couples: results from a nationwide randomized controlled trial. J Consult Clin Psychol 2020 Apr;88(4):283-294. [doi: 10.1037/ccp0000479] [Medline: 32134290]
4. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. J Community Health 2013 Oct;38(5):976-993 [FREE Full text] [doi: 10.1007/s10900-013-9681-1] [Medline: 23543372]
5. Talevi D, Socci V, Carai M, Carnaghi G, Faleri S, Trebbi E, et al. Mental health outcomes of the COVID-19 pandemic. Riv Psichiatr 2020;55(3):137-144. [doi: 10.1708/3382.33569] [Medline: 32489190]
6. McLean SA, Booth AT, Schnabel A, Wright BJ, Painter FL, McIntosh JE. Exploring the efficacy of telehealth for family therapy through systematic, meta-analytic, and qualitative evidence. Clin Child Fam Psychol Rev 2021 Jun;24(2):244-266 [FREE Full text] [doi: 10.1007/s10567-020-00340-2] [Medline: 33492545]
7. Florean IS, Dobrean A, Păsărelu CR, Georgescu RD, Milea I. The efficacy of internet-based parenting programs for children and adolescents with behavior problems: a meta-analysis of randomized clinical trials. Clin Child Fam Psychol Rev 2020 Dec 08;23(4):510-528. [doi: 10.1007/s10567-020-00326-0] [Medline: 32897527]
8. Baumel A, Correll CU, Birnbaum M. Adaptation of a peer based online emotional support program as an adjunct to treatment for people with schizophrenia-spectrum disorders. Internet Interv 2016 May;4:35-42 [FREE Full text] [doi: 10.1016/j.invent.2016.03.003] [Medline: 30135788]
9. Carr A. Couple therapy, family therapy and systemic interventions for adult‐focused problems: the current evidence base. J Fam Ther 2018 Jun 13;41(4):492-536. [doi: 10.1111/1467-6427.12225]
10. de Boer K, Muir SD, Silva SS, Nedeljkovic M, Seabrook E, Thomas N, et al. Videoconferencing psychotherapy for couples and families: a systematic review. J Marital Fam Ther 2021 Apr 10;47(2):259-288. [doi: 10.1111/jmft.12518] [Medline: 33837968]
11. Köhnen M, Dreier M, Seeralan T, Kriston L, Härter M, Baumeister H, et al. Evidence on technology-based psychological interventions in diagnosed depression: systematic review. JMIR Ment Health 2021 Feb 10;8(2):e21700 [FREE Full text] [doi: 10.2196/21700] [Medline: 33565981]
12. Sourander A, McGrath PJ, Ristkari T, Cunningham C, Huttunen J, Lingley-Pottie P, et al. Internet-assisted parent training intervention for disruptive behavior in 4-year-old children: a randomized clinical trial. JAMA Psychiatry 2016 Apr 01;73(4):378-387. [doi: 10.1001/jamapsychiatry.2015.3411] [Medline: 26913614]
13. von Sydow K, Retzlaff R, Beher S, Haun MW, Schweitzer J. The efficacy of systemic therapy for childhood and adolescent externalizing disorders: a systematic review of 47 RCT. Fam Process 2013 Dec 11;52(4):576-618. [doi: 10.1111/famp.12047] [Medline: 24102196]
14. von Sydow K, Retzlaff R. Aktueller stand der systemischen therapie. Psychotherapeut 2021 Oct 28;66(6):469-477. [doi: 10.1007/S00278-021-00547-W]
15. Retzlaff R, von Sydow K, Beher S, Haun MW, Schweitzer J. The efficacy of systemic therapy for internalizing and other disorders of childhood and adolescence: a systematic review of 38 randomized trials. Fam Process 2013 Dec 05;52(4):619-652. [doi: 10.1111/famp.12041] [Medline: 24329407]

XSL•FO
RenderX

16. Riedinger V, Pinquart M, Teubert D. Effects of systemic therapy on mental health of children and adolescents: a meta-analysis. J Clin Child Adolesc Psychol 2017 Oct 14;46(6):880-894. [doi: 10.1080/15374416.2015.1063427] [Medline: 26467300]
17. Stratton P, Lask J. The development of systemic family therapy for changing times in the United Kingdom. Contemp Fam Ther 2013 Mar 13;35(2):257-274. [doi: 10.1007/S10591-013-9252-8]
18. von Sydow K, Stefan Beher S, Retzlaff R, Schweitzer J. Wie wirksam ist Systemische Psychotherapie/Familientherapie? Psychiatr Prax 2008 Feb;8(1):20-27. [doi: 10.1026/1860-7357.8.1.20]
19. Allgemeine Methoden. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). URL: https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf [accessed 2024-04-29]
20. Borcsa M, Pomini V. La diffusione delle pratiche digitali nella psicoterapia sistemica: lo stato dell'arte in Europa e oltre. In: Manfrida G, Albertini V, Eisenberg E, editors. La clinica e il web. Risorse tecnologiche e comunicazione psicoterepeutica online. Milano, Italy: FrancoAngeli; Jul 07, 2020:37-60 URL: https://www.stateofmind.it/2020/09/clinica-web-recensione/
21. Lebow JL. Couple and family therapy (and families) in the digital age. Fam Process 2021 Sep 27;60(3):689-693. [doi: 10.1111/famp.12705] [Medline: 34580874]
22. Probst T, Stippl P, Pieh C. Changes in provision of psychotherapy in the early weeks of the COVID-19 lockdown in Austria. Int J Environ Res Public Health 2020 May 27;17(11):3815 [FREE Full text] [doi: 10.3390/ijerph17113815] [Medline: 32471295]
23. Burbach FA, Pote H. Digital approaches – a paradigm shift? J Fam Ther 2021 Mar 18;43(2):169-184. [doi: 10.1111/1467-6427.12336]
24. Borcsa M, Pomini V, Saint‐Mont U. Digital systemic practices in Europe: a survey before the COVID‐19 pandemic. J Fam Ther 2020 Sep 09;43(1):4-26. [doi: 10.1111/1467-6427.12308]
25. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
26. 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]
27. Pinsof W. Integrative systemische therapie (IST). In: von Sydow KB, Borst U, editors. Systemische Therapie in der Praxis. Berlin, Germany: Beltz, Weinheim; Nov 13, 2018:963-975.
28. von Sydow KB, Borst U. Systemische Therapie in der Praxis. Berlin, Germany: Beltz; Nov 13, 2018.
29. Duke DC, Wagner DV, Ulrich J, Freeman KA, Harris MA. Videoconferencing for teens with diabetes: family matters. J Diabetes Sci Technol 2016 Jul;10(4):816-823 [FREE Full text] [doi: 10.1177/1932296816642577] [Medline: 27075708]
30. Murry VM, Berkel C, Liu N. The closing digital divide: delivery modality and family attendance in the pathways for African American success (PAAS) program. Prev Sci 2018 Jul 22;19(5):642-651 [FREE Full text] [doi: 10.1007/s11121-018-0863-z] [Medline: 29356928]
31. Kurowski BG, Stancin T, Taylor HG, McNally KA, Kirkwood MW, Cassedy A, et al. Comparative effectiveness of family problem-solving therapy (F-PST) for adolescents after traumatic brain injury: Protocol for a randomized, multicenter, clinical trial. Contemp Clin Trials Commun 2018 Jun;10:111-120 [FREE Full text] [doi: 10.1016/j.conctc.2018.04.001] [Medline: 30023445]
32. Franta C, Philipp J, Waldherr K, Truttmann S, Merl E, Schöfbeck G, et al. Supporting carers of children and adolescents with eating disorders in Austria (SUCCEAT): study protocol for a randomised controlled trial. Eur Eat Disord Rev 2018 Sep 06;26(5):447-461 [FREE Full text] [doi: 10.1002/erv.2600] [Medline: 29732651]
33. Freeman KA, Duke DC, Harris MA. Behavioral health care for adolescents with poorly controlled diabetes via Skype: does working alliance remain intact? J Diabetes Sci Technol 2013 May 01;7(3):727-735 [FREE Full text] [doi: 10.1177/193229681300700318] [Medline: 23759406]
34. Bramer WM, de Jonge GB, Rethlefsen ML, Mast F, Kleijnen J. A systematic approach to searching: an efficient and complete method to develop literature searches. J Med Libr Assoc 2018 Oct 04;106(4):531-541 [FREE Full text] [doi: 10.5195/jmla.2018.283] [Medline: 30271302]
35. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Res Synth Methods 2018 Dec 07;9(4):602-614 [FREE Full text] [doi: 10.1002/jrsm.1287] [Medline: 29314757]
36. Thomas J, McDonald S, Noel-Storr A, Shemilt I, Elliott J, Mavergames C, et al. Machine learning reduced workload with minimal risk of missing studies: development and evaluation of a randomized controlled trial classifier for Cochrane Reviews. J Clin Epidemiol 2021 May;133:140-151 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.11.003] [Medline: 33171275]
37. Nussbaumer-Streit B, Mayr V, Dobrescu A, Chapman A, Persad E, Klerings I, et al. The effectiveness of quarantine alone or in combination with other public health measures to control coronavirus disease 2019: a rapid review. Cochrane Database Syst Rev 2020 Mar 27;9(Suppl 1):1 [FREE Full text] [doi: 10.1002/14651858.CD202001]
38. Hefti S, Pérez T, Fürstenau U, Rhiner B, Swenson CC, Schmid M. Multisystemic therapy for child abuse and neglect: do parents show improvement in parental mental health problems and parental stress? J Marital Fam Ther 2020 Jan 05;46(1):95-109. [doi: 10.1111/jmft.12367] [Medline: 30516844]

XSL•FO
RenderX

39. Tomm K. Introducing the IPscope: a systemic assessment tool for distinguishing interpersonal patterns. In: Tomm K, St. George S, Wulff D, Strong T, editors. Patterns in Interpersonal Interactions: Inviting Relational Understandings for Therapeutic Change. New York, NY: Routledge; 2014:13-35.
40. Ataoglu A, Ozcetin A, Icmeli C, Ozbulut O. Paradoxical therapy in conversion reaction. J Korean Med Sci 2003 Aug;18(4):581-584 [FREE Full text] [doi: 10.3346/jkms.2003.18.4.581] [Medline: 12923337]
41. Pinsof W, Breunlin D, Russell W, Lebow J, Chambers A. Integrative Systemic Therapy: Metaframeworks for Problem Solving With Individuals, Couples, and Families. New York, NY: American Psychological Association; 2018.
42. Hale D, Frusha C. MRI brief therapy: a tried and true systemic approach. J Syst Ther 2016 Jun;35(2):14-24. [doi: 10.1521/jsyt.2016.35.2.14]
43. Vitry G, Pakrosnis R, Brosseau OG, Duriez N. Effectiveness and efficiency of strategic and systemic therapy in naturalistic settings: preliminary results from a systemic practice research network (SYPRENE). J Fam Ther 2021 Apr 16;43(4):516-537. [doi: 10.1111/1467-6427.12343]
44. Smoktunowicz E, Barak A, Andersson G, Banos RM, Berger T, Botella C, et al. Consensus statement on the problem of terminology in psychological interventions using the internet or digital components. Internet Interv 2020 Sep;21:100331 [FREE Full text] [doi: 10.1016/j.invent.2020.100331] [Medline: 32577404]
45. Knopp K, Schnitzer JS, Khalifian C, Grubbs K, Morland LA, Depp C. Digital interventions for couples: state of the field and future directions. Couple Family Psychol Res Pract 2023 Dec;12(4):201-217. [doi: 10.1037/cfp0000213]
46. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
47. Helps S, Le Coyte Grinney M. Synchronous digital couple and family psychotherapy: a meta‐narrative review. J Fam Ther 2021 Jan 27;43(2):185-214. [doi: 10.1111/1467-6427.12333]
48. Baumel A, Pawar A, Kane JM, Correll CU. Digital parent training for children with disruptive behaviors: systematic review and meta-analysis of randomized trials. J Child Adolesc Psychopharmacol 2016 Oct;26(8):740-749. [doi: 10.1089/cap.2016.0048] [Medline: 27286325]
49. Leijten P, Gardner F, Melendez-Torres GJ, van Aar J, Hutchings J, Schulz S, et al. Meta-analyses: key parenting program components for disruptive child behavior. J Am Acad Child Adolesc Psychiatry 2019 Feb;58(2):180-190. [doi: 10.1016/j.jaac.2018.07.900] [Medline: 30738545]
50. Carr A. Family therapy and systemic interventions for child‐focused problems: the current evidence base. J Fam Ther 2018 Apr 27;41(2):153-213. [doi: 10.1111/1467-6427.12226]
51. Gelin Z, Cook-Darzens S, Simon Y, Hendrick S. Two models of multiple family therapy in the treatment of adolescent anorexia nervosa: a systematic review. Eat Weight Disord 2016 Mar;21(1):19-30. [doi: 10.1007/s40519-015-0207-y] [Medline: 26223191]
52. Pinquart M, Oslejsek B, Teubert D. Efficacy of systemic therapy on adults with mental disorders: a meta-analysis. Psychother Res 2016;26(2):241-257. [doi: 10.1080/10503307.2014.935830] [Medline: 25032487]
53. Köhnen M, Kriston L, Härter M, Dirmaier J, Liebherz S. Rationale and design of a systematic review: effectiveness and acceptance of technology-based psychological interventions in different clinical phases of depression management. BMJ Open 2019 Mar 27;9(3):e028042 [FREE Full text] [doi: 10.1136/bmjopen-2018-028042] [Medline: 30918040]
54. Hair K, Bahor Z, Macleod M, Liao J, Sena E. The automated systematic search deduplicator (ASySD): a rapid, open-source, interoperable tool to remove duplicate citations in biomedical systematic reviews. BMC Biol 2023 Sep 07;21(1):189 [FREE Full text] [doi: 10.1186/s12915-023-01686-z] [Medline: 37674179]
55. Higgins JP, Savović J, Page MJ, Elbers RS, Sterne JA. Assessing risk of bias in a randomized trial. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, NY: The Cochrane Collaboration; 2021.
56. Murry VM, Berkel C, Inniss-Thompson MN, Debreaux ML. Pathways for African American success: results of three-arm randomized trial to test the effects of technology-based delivery for rural African American families. J Pediatr Psychol 2019 Apr 01;44(3):375-387 [FREE Full text] [doi: 10.1093/jpepsy/jsz001] [Medline: 30865782]
57. Riley AR, Duke DC, Freeman KA, Hood KK, Harris MA. Depressive symptoms in a trial behavioral family systems therapy for diabetes: a post hoc analysis of change. Diabetes Care 2015 Aug;38(8):1435-1440. [doi: 10.2337/dc14-2519] [Medline: 26015558]
58. Harris MA, Freeman KA, Duke DC. Seeing is believing: using Skype to improve diabetes outcomes in youth. Diabetes Care 2015 Aug;38(8):1427-1434. [doi: 10.2337/dc14-2469] [Medline: 26033508]
59. Murry VM, Kettrey HH, Berkel C, Inniss-Thompson MN. The pathways for African American success: does delivery platform matter in the prevention of HIV risk vulnerability among youth? J Adolesc Health 2019 Aug;65(2):255-261. [doi: 10.1016/j.jadohealth.2019.02.013] [Medline: 31043346]
60. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
61. McGuinness LA, Higgins JP. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan 06;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]

XSL•FO
RenderX

62. Drahota A, Beller E. RevMan calculator. The Cochrane Collaboration. URL: https://training.cochrane.org/resource/revman-calculator [accessed 2024-04-29]
63. Deeks JJ, Higgins JP, Altman DG, McKenzie JE, Veroniki AA. Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions: Cochrane Book Series. New York, NY: The Cochrane Collaboration; 2022.
64. Truttmann S, Philipp J, Zeiler M, Franta C, Wittek T, Merl E, et al. Long-term efficacy of the workshop vs. online SUCCEAT (supporting carers of children and adolescents with eating disorders) intervention for parents: a quasi-randomised feasibility trial. J Clin Med 2020 Jun 18;9(6):1912 [FREE Full text] [doi: 10.3390/jcm9061912] [Medline: 32570930]
65. Kurowski BG, Taylor HG, McNally KA, Kirkwood MW, Cassedy A, Horn PS, et al. Online family problem-solving therapy (F-PST) for executive and behavioral dysfunction after traumatic brain injury in adolescents: a randomized, multicenter, comparative effectiveness clinical trial. J Head Trauma Rehabil 2020;35(3):165-174 [FREE Full text] [doi: 10.1097/HTR.0000000000000545] [Medline: 31834062]
66. Wade SL, Cassedy AE, McNally KA, Kurowski BG, Kirkwood MW, Stancin T, et al. A randomized comparative effectiveness trial of family-problem-solving treatment for adolescent brain injury: parent outcomes from the coping with head injury through problem solving (CHIPS) study. J Head Trauma Rehabil 2019;34(6):E1-E9. [doi: 10.1097/HTR.0000000000000487] [Medline: 31033747]
67. Wade SL, Cassedy AE, Sklut M, Taylor HG, McNally KA, Kirkwood MW, et al. The relationship of adolescent and parent preferences for treatment modality with satisfaction, attrition, adherence, and efficacy: the coping with head injury through problem-solving (CHIPS) study. J Pediatr Psychol 2019 Apr 01;44(3):388-401. [doi: 10.1093/jpepsy/jsy087] [Medline: 30452665]
68. Wade SL, Cassedy AE, Taylor HG, McNally KA, Kirkwood MW, Stancin T, et al. Adolescent quality of life following family problem-solving treatment for brain injury. J Consult Clin Psychol 2019 Nov;87(11):1043-1055. [doi: 10.1037/ccp0000440] [Medline: 31556652]
69. Greenwood H, Krzyzaniak N, Peiris R, Clark J, Scott AM, Cardona M, et al. Telehealth versus face-to-face psychotherapy for less common mental health conditions: systematic review and meta-analysis of randomized controlled trials. JMIR Ment Health 2022 Mar 11;9(3):e31780 [FREE Full text] [doi: 10.2196/31780] [Medline: 35275081]
70. Page MJ, Higgins JP, Sterne JA. Assessing risk of bias due to missing results in a synthesis. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. New York, NY: The Cochrane Collaboration; 2022.
71. Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. Psychother Res 2018 May 09;28(3):347-355. [doi: 10.1080/10503307.2017.1411628] [Medline: 29224503]
72. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. J Educ Behave Stat 2005 Sep 01;30(3):261-293. [doi: 10.3102/10769986030003261]
73. McAloon J, de la Poer Beresford K. Online behavioral parenting interventions for disruptive behavioral disorders: a PRISMA based systematic review of clinical trials. Child Psychiatry Hum Dev 2023 Apr 24;54(2):379-396. [doi: 10.1007/s10578-021-01253-z] [Medline: 34561755]
74. Boston P. Systemic family therapy and the influence of post-modernism. Adv Psychiatr Treat 2018 Jan 02;6(6):450-457. [doi: 10.1192/apt.6.6.450]
75. Lorås L, Bertrando P, Ness O. Researching systemic therapy history: in search of a definition. J Fam Psychother 2017 Feb 17;28(2):134-149. [doi: 10.1080/08975353.2017.1285656]
76. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ 2020 Jun 04;369:m1714 [FREE Full text] [doi: 10.1136/bmj.m1714] [Medline: 32499297]
77. Hasselberg MJ. The digital revolution in behavioral health. J Am Psychiatr Nurses Assoc 2020 Oct 05;26(1):102-111. [doi: 10.1177/1078390319879750] [Medline: 31587624]
78. Greene CJ, Morland LA, Durkalski VL, Frueh BC. Noninferiority and equivalence designs: issues and implications for mental health research. J Trauma Stress 2008 Oct;21(5):433-439 [FREE Full text] [doi: 10.1002/jts.20367] [Medline: 18956449]
79. 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 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]

Abbreviations

BFST-D: Behavioral Family Systems Therapy for Diabetes
CPP: constituent primary part
F-PST: Family Problem-Solving Therapy
MD: mean difference
PAAS: Pathways for African American Success

XSL•FO
RenderX

PRISMA: Preferred Reporting Item for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB: risk of bias
ST: systemic therapy
SUCCEAT: Supporting Carers of Children and Adolescents with Eating Disorders in Austria

Edited by T de Azevedo Cardoso; submitted 07.03.23; peer-reviewed by M Rollwage, R Halkes; comments to author 01.06.23; revised version received 18.07.23; accepted 08.08.24; published 24.02.25.

Please cite as:
Erasmus P, Borrmann M, Becker J, Kuchinke L, Meinlschmidt G
Comparing Digital Versus Face-to-Face Delivery of Systemic Psychotherapy Interventions: Systematic Review and Meta-Analysis of Randomized Controlled Trials
Interact J Med Res 2025;14:e46441
URL: https://www.i-jmr.org/2025/1/e46441
doi:10.2196/46441
PMID:

©Pieter Erasmus, Moritz Borrmann, Jule Becker, Lars Kuchinke, Gunther Meinlschmidt. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 24.02.2025. 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

Global Evidence on the Sustainability of Telemedicine in Outpatient and Primary Care During the First 2 Years of the COVID-19 Pandemic: Scoping Review Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) Framework

Daniela Valdes[1], MSc; Ankit Shanker[2], MSc; Ghofran Hijazi[1], MSc; Daniel Opoku Mensah[2], PhD; Tahir Bockarie[2], PhD; Ioana Lazar[3], MSc; Siti Aishah Ibrahim[2], MSc; Hamid Zolfagharinia[4], MSc; Rob Procter[1,5], PhD; Rachel Spencer[2], MD, PhD; Jeremy Dale[2], MA, MB, PhD; Armina Paule[2], MSc; Liam Jonathon Medlin[6], BSc; Keerthana Tharuvara Kallottil[3], MSc

[1]Department of Computer Science, University of Warwick, Coventry, United Kingdom

[2]Warwick Medical School, University of Warwick, Coventry, United Kingdom

[3]Warwick Manufacturing Group, University of Warwick, Coventry, United Kingdom

[4]Research & Innovation, Birmingham Community Healthcare Foundation Trust, National Health Service, Birmingham, United Kingdom

[5]The Alan Turing Institute for Data Science and AI, London, United Kingdom

[6]Department of Chemistry, University of Warwick, Coventry, United Kingdom

Corresponding Author:
Daniela Valdes, MSc
Department of Computer Science
University of Warwick
6 Lord Bhattacharyya Way
Coventry, CV4 7AL
United Kingdom
Phone: 44 7943058485
Email: daniela.valdes@warwick.ac.uk

Abstract

Background: The rapid implementation of telemedicine during the early stages of the COVID-19 pandemic raises questions about the sustainability of this intervention at the global level.

Objective: This research examines the patient experience, health inequalities, and clinician-patient relationship in telemedicine during the COVID-19 pandemic's first 2 years, aiming to identify sustainability factors.

Methods: This study was based on a prepublished protocol using the Joanna Briggs Institute (JBI) methodology for scoping reviews. We included academic and gray literature published between March 2020 and March 2022 according to these criteria: (1) population (any group); (2) concepts (patient experience, clinician-patient relationship, health inequalities); (3) context (telemedicine in primary and outpatient care); (4) excluding studies pertaining to surgery, oncology, and (inpatient) psychiatry. We searched Ovid Medline/PubMed (January 1, 2022), Web of Science (March 19, 2022), Google/Google Scholar (February and March 2022), and others. The risk of bias was not assessed as per guidance. We used an analysis table for the studies and color-coded tabular mapping against a health care technology adoption framework to identify sustainability (using double-blind extraction).

Results: Of the 134 studies that met our criteria, 49.3% (66/134) reported no specific population group. Regarding the concepts, 41.8% (56/134) combined 2 of the concepts studied. The context analysis identified that 56.0% (75/134) of the studies referred to, according to the definition in the United Kingdom, an outpatient (ambulatory care) setting, and 34.3% (46/134) referred to primary care. The patient experience analysis reflected positive satisfaction and sustained access during lockdowns. The clinician-patient relationship impacts were nuanced, affecting interaction and encounter quality. When mapping to the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework, 81.3% (109/134) of the studies referenced the innovation's

XSL•FO
RenderX

sustainability. Although positive overall, there were some concerns about sustainability based on quality, eHealth literacy, and access to health care for vulnerable migrants and the uninsured.

Conclusions: We identified confusion between the concepts of patient experience and patient satisfaction; therefore, future research could focus on established frameworks to qualify the patient experience across the whole pathway and not just the remote encounter. As expected, our research found mainly descriptive analyses, so there is a need for more robust evidence methods identifying impacts of changes in treatment pathways. This study illustrates modern methods to decolonize academic research by using gray literature extracts in other languages. We acknowledge that the use of Google to identify gray literature at the global level and in other languages has implications on reproducibility. We did not consider synchronous text-based communication.

Trial Registration: Open Science Framework 4z5ut; https://osf.io/4z5ut/

(Interact J Med Res 2025;14:e45367) doi:10.2196/45367

KEYWORDS

pandemic; primary care; outpatients; telemedicine; ambulatory care; global health; patient experience; NASSS; clinician-patient relationship; health inequalities; gray literature; PRISMA

Introduction

Following the World Health Organization (WHO) announcement declaring COVID-19 a pandemic on March 11, 2020, the organization recommended telemedicine as one of the first critical interventions to minimize demands on stretched supplies of personal protective equipment [1,2]. Although telehealth was not new as a delivery mode, there were great expectations particularly around the use of video consultations in this context. In Africa, telemedicine held promise, as it rationalized human resources allowing national or international experts to relay advice to other clinicians [3]; in the United Kingdom, the pandemic was deemed a “burning platform” to propel the UK National Health Service (NHS) toward widespread adoption of video consultations [4]. Most medical specialties responded to the WHO recommendation with rapid changes in service delivery toward telemedicine (both telephone and a new medium—video consultations) in primary and secondary care across the globe.

In the United Kingdom, 15 months into the pandemic, the academic community and political groups raised questions around the sustainability and impacts of the move toward telemedicine, building upon the learning of the past year toward a “new normal,” particularly as social distancing and lockdown measures were removed [5]. A 2021 report by the UK Health and Social Care Committee defined telemedicine as a “welcomed and positive innovation” overall while highlighting concerns by various national organizations on its impact on health inequalities in terms of exclusion (lack of access), quality, and patient safety. The committee reported important consensus from recognized local institutions such as the Health Foundation and the Kings Fund and patient organizations such as National Voices and Healthwatch on the need for additional research to assess the future of telemedicine in a patient-centered way [6].

In previous protocols [7,8], we relayed how prepandemic evidence synthesis identified several barriers and objections that hindered telemedicine uptake, including technology, workload, and confidentiality [9], as well as concerns regarding appropriateness [10]. However, at the outset of the pandemic, these objections were rapidly overcome, supported by major regulatory and financial enablers [11,12]. Given the considerable incentives and support toward the implementation of telemedicine, there are concerns about the risks of losing some of the advantages of this mode of delivery in a postpandemic future [13], particularly once incentives are no longer in place [14,15]. These concerns apply to not only the United Kingdom, Canada, or the United States but also sub-Saharan Africa [16] and Latin America [17] where considerable barriers persist and there has been more limited evidence of uptake.

The purpose of this scoping review was to explore the global evidence (both academic and nonacademic) surrounding the rapid adoption of telemedicine in outpatient and primary care settings during the first 2 years of the COVID-19 pandemic to identify how elements related to patient experience, clinician-patient relationship, and health inequalities support (or take away from) the sustainability of this delivery model.

Methods

Protocol

This review was conducted according to an a priori published protocol [8] following the Joanna Briggs Institute (JBI) methodology for scoping reviews, updated guidance, and data extraction guidance [18-20]. This last guidance clarifies that a review can focus on the most relevant section of a document for analysis, without having to review whole studies in scoping reviews [20]. We outlined deviations from the original protocol and the methodology in Multimedia Appendix 1. A key contribution of this study hinges on the methods used to search and extract gray literature across a wider set of countries as to achieve truly global representation.

Inclusion Criteria

Population, Concept, Context Principle

The inclusion criteria used the Population, Concept, Context (PCC) principle. We classified each article against the PCC framework. Namely, for each document, we sought to identify (1) the population group (if any) to which the document referred, (2) the concept to which it referred (patient experience, clinician-patient relationship, health inequalities, or a combination of these), and (3) in which (clinical) context telemedicine was being used (outpatient, primary care, or particular specialties).

XSL•FO
RenderX

Population

The review focused on primary care services offered to the general population. Studies focusing on specific population groups or those with particular conditions within a particular country or geographical area were included.

Concept

Although the key concept under consideration was the adoption of "telemedicine," as defined in the Introduction, we narrowed our inclusion criteria to the sustainability of the interventions, focusing on patient experience, health inequalities, and clinician-patient relationship. Telemedicine has been defined as per our protocol [8] using academic literature [21] and the WHO [22], which defines it as "The delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies [...]." Although we focused on telephone and video as communication technologies, we thank an anonymous peer reviewer who identified that text messaging also falls in this category (albeit not included in our review).

Context

The context was primary care services provided during the COVID-19 pandemic in any setting or country during the first 2 years of the pandemic (March 2020-March 2022).

Types of Sources

This scoping review considered quantitative, qualitative, and mixed methods study designs for inclusion. In addition, systematic reviews, protocols, other documents, and commentaries or opinions were considered for inclusion in the proposed scoping review. These commentaries or other documents might have appeared in peer-reviewed journals or other gray literature such as industry magazines or reports [23,24].

Search Strategy

This section summarizes our prepublished protocol [8]. We structured this section by first explaining the selection and identification of search terms (both in English and other languages), then how we used those terms to search for academic and gray literature. To note, the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [25] was used to structure our review protocol [8].

Identifying Search Terms

Our search terms, which were in English, centered around telemedicine, primary care, and COVID-19 and were expanded through a limited search on Ovid Medline and Web of Science. We also sought guidance from University of Warwick librarians, who identified COVID-19 search terms from the National Institute for Health and Care Excellence (NICE) [26]. Additionally, we examined previous telemedicine protocols in general practice [27] and Primary Care Cochrane Library Protocols. For search terms in other languages, we used a 2-step process to identify relevant terms for the included non-English search results. We used Google Translate and engaged in discussions with native speakers to ensure these were accurate translations.

Search Approaches

We had 2 distinct approaches depending on whether we were searching for academic or gray literature. With the aforementioned terms, we searched the following academic databases: Ovid Medline (equivalent to PubMed [28]), Web of Science, and Google Scholar. PROSPERO and Cochrane Library were used to inform our design. PROSPERO in particular was used to check whether there was any ongoing review on the topic. All identified search terms and examples of the searches (for academic and gray literature) can be found in Multimedia Appendix 2, as per our previously published protocol [8].

We searched Ovid Medline/PubMed (January 13, 2022), Web of Science (March 19, 2022), Livivo (March 15, 2022), Scopus (March 19, 2022), PROSPERO (January 12, 2022), Cochrane Library (January 12, 2022), and Google/Google Scholar (February 2022 and March 2022).

As shown in Multimedia Appendix 2, to identify relevant gray literature published at the time of the pandemic, we used advanced Google search criteria with simplified search terms. We combined the terms telemedicine, "Primary Care," and COVID-19 with (1) "patient experience," (2) "health inequalities," or (3) "patient-clinician interaction" and asked the search engine to provide pdf-only results within the years 2020 through 2022. Although we recognize it is not fully possible to reproduce Google searches, the selection of pdf documents was aimed at identifying the most retrievable and credible gray literature [8] while, at the same time, supporting reproducibility of the analysis [23,24].

We undertook these searches in English, Chinese, Spanish, Arabic, Portuguese, Hindi, and Indonesian. To improve representation of African countries due to difficulties searching in Urdu, we undertook additional Google searches in English including country-specific results for the 5 largest African countries by population (Nigeria, Ethiopia, Democratic Republic of Congo, Egypt, and South Africa [29]). For Google searches, we selected the first 30 results by relevance, and for English-based results, we selected the top 10 results by country. We selected the first 30 results based on the literature [30-33], timeline, and budget. Further, as can be seen in the example for a failed search for Pakistan in Multimedia Appendix 2, Google searches only provided less than a handful of results when restricting by country of publication.

We followed established guidelines for analyzing non-English text [34]. We used machine translation via Google Translate to translate at least 3 paragraphs containing the key search terms (telemedicine, primary care, and any of the combinations of patient experience, clinician-patient relationship, and health inequalities). We selected 3 paragraphs that were roughly equivalent in the number of words to that of an abstract.

Screening

The authors used a single-phase, double-blind screening of abstracts and extracts based on the eligibility criteria. Screening instructions were tested by 4 authors across a sample of 50

XSL•FO
RenderX

abstracts to verify the instructions had been properly understood. The remaining articles were allocated across several combinations of pairs of authors using double-blind screening and Rayyan [35] as an aid. Any discrepancies were resolved by the pair of authors themselves and verified by the lead author before data extraction. Given the extended scope of our review, and in agreement with the prepublished protocol and the latest JBI guidance [20], there was no full-text screening as the abstracts and extracts were our main data source. Abstracts and extracts have been made available in Multimedia Appendix 3.

Data Extraction

The authors used double-blind extraction of data using an Excel table template as outlined in Multimedia Appendix 4. Following the latest JBI guidance for scoping reviews [20], we chose to focus data extraction on abstracts and extracts only, reflecting the wide research design that allowed us to accommodate the breadth of the review in terms of (1) themes; (2) a worldwide, multilanguage approach; and (3) sources. Using a data extraction tool shared in Multimedia Appendix 4, we mapped the text in tabular form against the NASSS framework [36] domains, noting some of the document sections might touch upon 1 or more domains.

Data Analysis and Presentation

In agreement with the latest JBI methodological guidance [18-20], no critical appraisal was undertaken, and the final presentation of results consisted of the following. In the first section, the results of the search were presented in a PRISMA-ScR flow diagram [25], including a flowchart and a checklist, and we conducted a table analysis of more detailed characteristics of the included documents (Table S1 in Multimedia Appendix 5). We captured the number of articles that included certain data against the PCC framework. Understandably, in the “population” label, the numbers do not add up to the total as some categories are not mutually exclusive. A color-coded (heat map) mapping was created in tabular form against the NASSS framework’s [36] domains: (1) the condition, (2) the technology, (3) the value proposition, (4) adopters, (5) organizations, (6) wider system, and (7) embedding and adaptation over time. The heat map shows graphically the maximum and minimum numbers of references for each domain (using the average counts for double-blind data extraction and mapping). We performed a narrative analysis (including sentiment analysis) of references to the sustainability of video consultations outside of social distancing restrictions brought about by the pandemic. The results are discussed from the point of view of sociotechnical grounded theory, providing strengths and limitations of the sources and the review method itself. We included a statement of positionality in our conclusions, as well as recommendations for research and practice. We believe that reflexivity through researcher positionality is fundamental to decolonizing global health research that seeks to include voices and perspectives usually marginalized from the academic discourse [37-39].

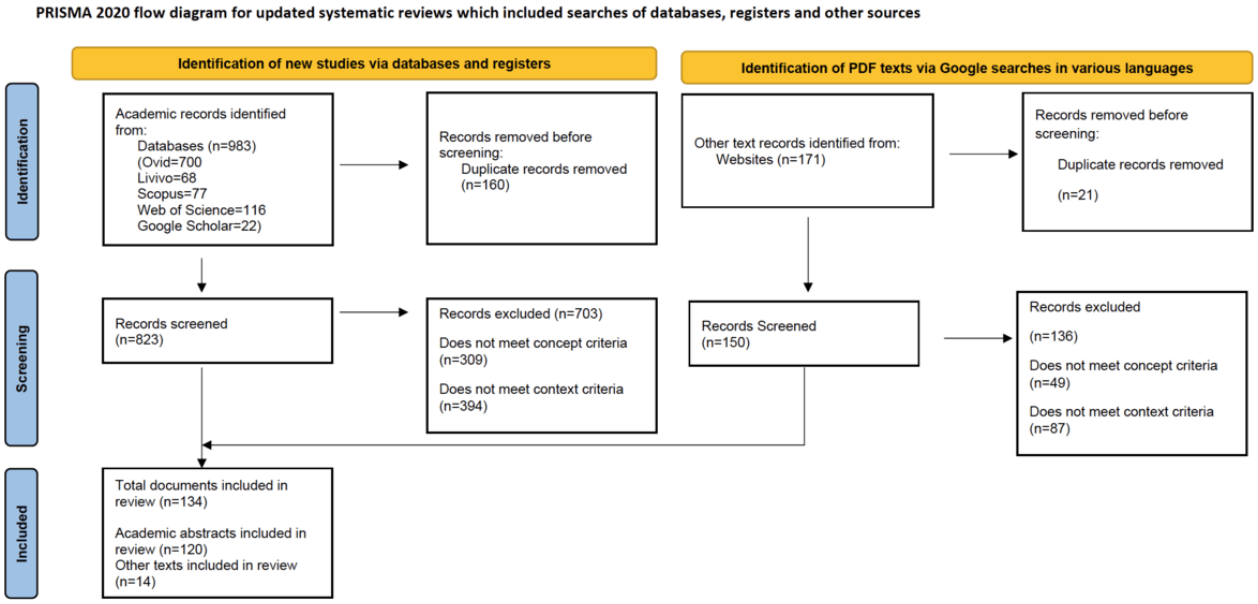
Results

PRISMA Study Inclusion

Academic database searches identified 983 records, and 171 additional records were identified through Google searches for various languages and countries. For Google Arabic and Google Pakistan, there were no results that matched the search criteria. Of the total 1154 articles, 181 were identified as duplicates and were removed: 160 duplicates were academic articles, and 21 duplicates were identified among the nonacademic articles. After the removal of duplicates, there were 973 articles to be screened. Of these 973 articles, 823 articles were from academic databases, and 150 were from nonacademic databases. After the abstract screening and before data extraction, 703 academic and 136 nonacademic articles were excluded, leaving a total of 134 documents [29,40-171] for data extraction including 120 academic articles and 14 extracts from gray literature searches. See the PRISMA-ScR [25] flowchart (Figure 1).

XSL•FO
RenderX

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flowchart outlining the process of identification, screening, and final inclusion across various types of data sources.



Summary of Document Types

We separated the academic studies (which we identified as those having a clearly labelled abstract with introduction, aims, methods, results, and discussion sections) from the other documents (or gray literature) and found that these other documents represented 27.6% (37/134) of the selected documents. Searches in academic databases and journals provided a small group of gray literature in the form of commentaries or guidelines. Conversely, general internet searches also identified a minority of academic documents.

The document pool achieved global representation, with documents from all continents, including South America [46,110,166] and Africa [116,121]. There was, however, overrepresentation from North America (58/134, 43%), with a large proportion of documents from the United States.

In terms of the methodology in the selected documents, over 72% (96/134) of the documents used surveys, questionnaires, or interviews. The use of surveys and questionnaires was closely related to the type of design observed, with most studies being cross-sectional (96/134, 72%). Finally, regarding the telemedicine medium, the documents did not specify the medium by generally referring to “telemedicine” in 57% (77/134) of documents. More details are provided in Table S2 in Multimedia Appendix 5.

Quantitative Classification Against the PCC Framework

Table S3 in Multimedia Appendix 5 includes the characteristics of the 134 documents against the protocol’s PCC. We follow with a short commentary highlighting any documents that exemplify these findings. Table 1 shows how the documents related to the various key concepts explored (clinician-patient relationship, health inequalities, and patient experience), and the following paragraphs summarize the key findings. In addition, Figure 2 shows the results of the bibliographic keyword analysis.

In terms of population groups, a large subset of abstracts and extracts (66/134, 49.3%) reported no specific demographics nor patient characteristics (see, for example, the extract from the board report from the East Kent Hospitals NHS Foundation Trust [64] or Karacabeyli et al [95]). The main demographic characteristics reported in abstracts and extracts were age (39/134, 29.1%) and sex (20/134, 14.9%), with studies also considering sociodemographic factors. It is important to note that a particular document might have tracked more than one characteristic; see, for example, Manski-Nankervis et al [113], which tracked education status, gender, age, and whether patients spoke English at home.

In terms of the concepts studied**,** the most popular concept was the clinician-patient relationship, reported in 28 abstracts. The majority of documents (69/134, 51.5%) combined 2 or 3 concepts. In the following paragraphs, we provide a brief summary of results across the various concepts.

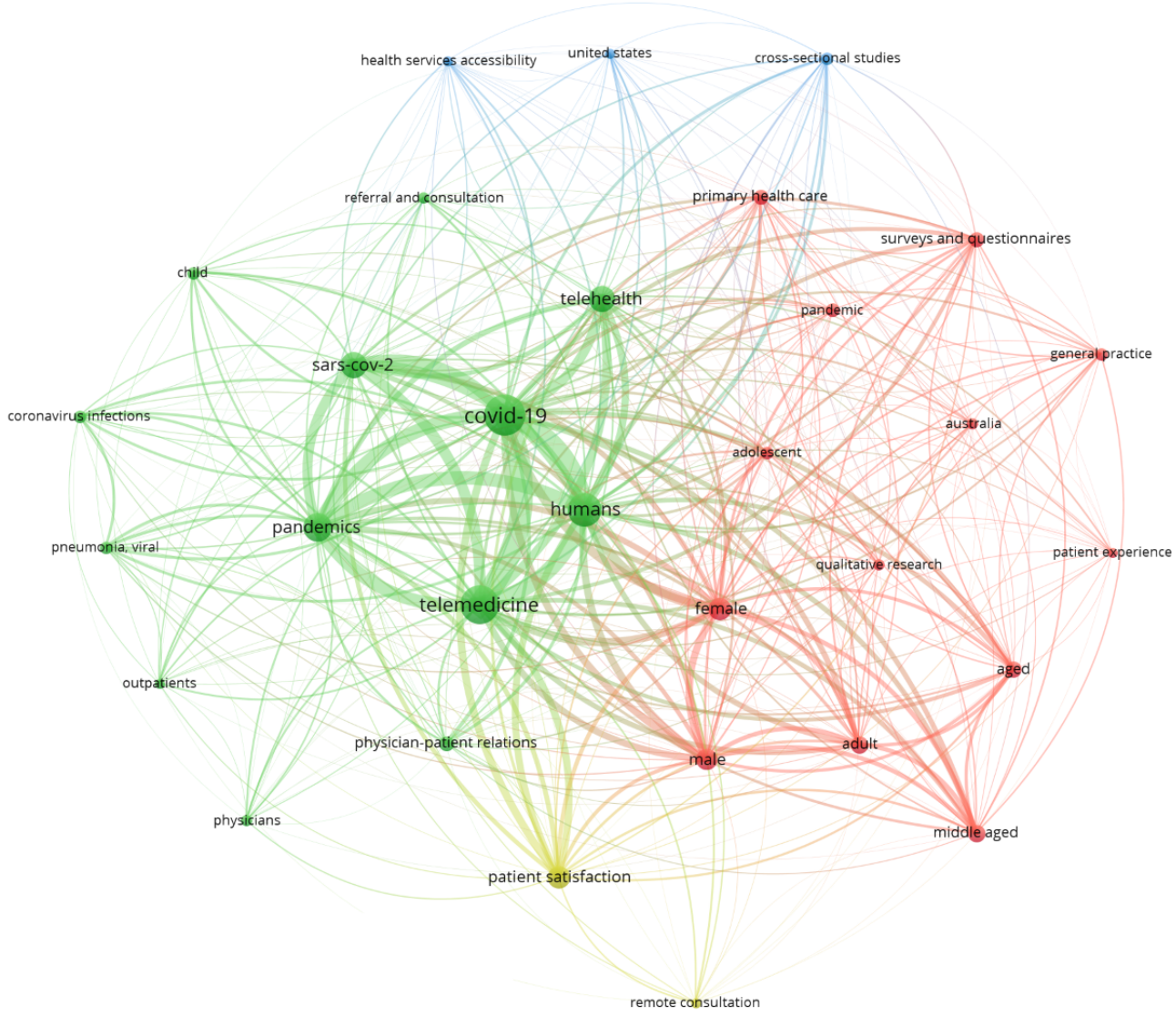
Regarding patient experience, we found 26 studies solely focused on this concept. Patient experience was mostly equated with patient satisfaction and access. There were positive levels of satisfaction overall [99,121,149] and sustained access at the time of lockdowns [48,53].

XSL•FO
RenderX

Table 1. Classification of the documents against the key concepts.

Concept	Documents, n
1 concept	
Clinician-patient relationship	28
Health inequalities	11
Patient experience	26
2 concepts	
Clinician-patient relationship, health inequalities	11
Clinician-patient relationship, patient experience	29
Health inequalities, patient experience	16
All 3 concepts	
Clinician-patient relationship, patient experience, health inequalities	13

Figure 2. Bibliographic keyword analysis.



Regarding health inequalities, there were only 11 studies solely focused on this concept. We found gaps in evidence that made it difficult to pinpoint the impacts on health inequalities of specific groups, with some evidence of health inequality exacerbation for those considered to be at the fringe of the economic and health system (such as vulnerable migrants, families with high levels of economic deprivation or health care insurance exclusion, or rural communities with limited access to technology) [74,116,140].

XSL•FO
RenderX

Regarding the clinician-patient relationship, there were 28 documents solely focused on this concept, mainly on the clinician-patient interaction and quality of the encounter. There was no consensus regarding whether the impact of telemedicine on the interaction had been positive, neutral, or negative [43,56,70], while there was consensus on the added value of the quality of the remote encounter for triage, follow-up consultations, or chronic condition management [128,133,155].

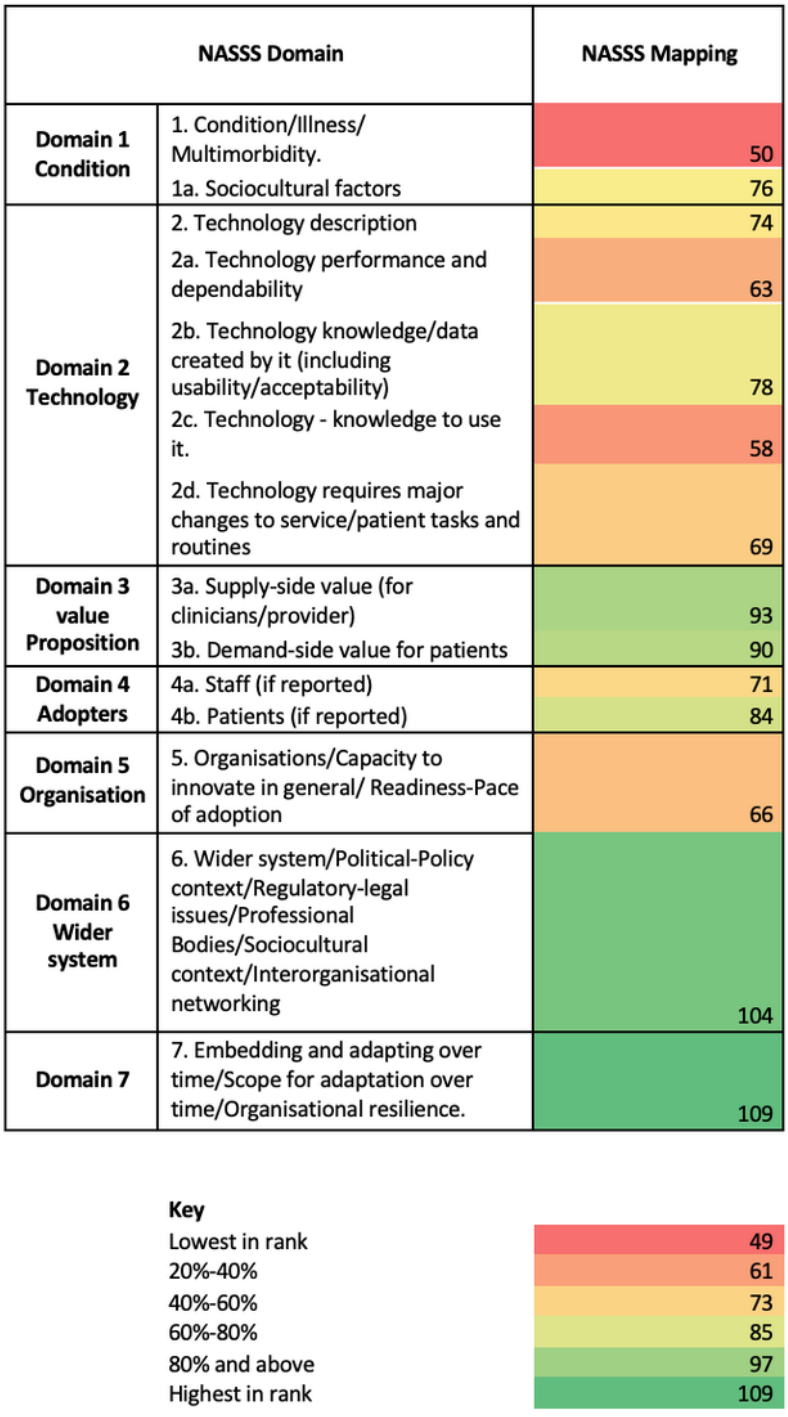
Finally, concerning the context, we identified 46 documents focused solely on primary care, general outpatient care featured in 19 documents, and various outpatient specialties were featured in 56 extracts.

Mapping Against the NASSS Framework

As shown in Figure 3, there was considerable variability in evidence across the NASSS [36] domains and subdomains. Domains 6 (wider system), 7 (time domain), and 3 (value proposition) had the most information. For domain 6, documents referenced the pandemic or pandemic-related lockdowns and infection control measures coupled with regulatory enablers or recommendations (such as mandatory online triage in primary care, parity in payments between face-to-face and telemedicine appointments). There was lower density of information about specific conditions that were being managed with telemedicine (subdomain 1a), knowledge needed to use telemedicine (subdomain 2c), and organizational or implementation aspects (domain 5).

XSL•FO
RenderX

Figure 3. Heat map against each of the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) domains.



Narrative Analysis of Sustainability

Roughly over one-half of the documents (75/134, 56%) identified challenges in terms of sustainability or made recommendations on how to address them. Among these, sentiment was mixed or neutral in 40 documents and positive in 31 documents, with only a small subgroup [68,89,141,148] viewing such challenges negatively as barriers to further planning and progress. Challenges and areas of consideration included general planning such as workforce training, digital resources, patient experience, and ethical issues [92] or a more focused look at the technology itself such as more effective digital platforms and increased use of home medical equipment [43]. Other issues such as patient selection were also noted with consideration to disease progression, language and cognitive ability, health literacy, and technology access [40]. The systematic review by Mbunge et al [116] of the digital technologies deployed in South Africa during the pandemic

XSL•FO
RenderX

identified a "digital divide" barrier in rural areas and advocated for better networks.

Almost one-third (38/134, 28%) of the documents set out plans for sustainable growth and further embedding, with generally positive sentiment. The report by the East Kent Hospitals NHS Foundation Trust in 2021 [64] mentioned an enhanced engagement plan to meet ambitious targets set by health care authorities for the delivery of telemedicine appointments. Tulupova et al [162] referred to plans for the creation of telemedicine guidelines and an educational program on communications in health care using digital technologies for patients to improve digital health literacy. The remaining documents either did not address the area of sustainability [42,44,49,66,95,101,106,110,111,122,132,143,153,167] or had a generally negative view on the sustainability of the intervention [63,74,80,112,169]. The negative commentary was based on concerns about quality (ie, treating musculoskeletal conditions), eHealth literacy, and access to health care for those at the fringe of the health care system coverage (such as vulnerable migrants and the uninsured).

Discussion

This section includes 4 main areas: (1) a summary of our findings, (2) a sociotechnical grounded theory research interpretation of the findings (based on Hoda [172]), (3) positioning against the wider and recent literature, and (4) strengths and limitations of our study. The section finalizes with a short conclusion.

Summary of Findings

Concerning the protocol's PCC, 49% (66/134) of the documents reported no specific population group targeted by the study (population); according to the UK definition, 55% (74/134) of the studies referred to an outpatient (ambulatory care) setting, and 34% (46/134) referred to primary care (context); and 49% (66/134) of the studies referred to only 1 of the concepts studied (concept).

Mapping to the NASSS framework [36] identified that 93% (125/134) of the studies referenced the sustainability of telemedicine with moderately positive comments.

Global Representation

We found 134 studies meeting our criteria and achieved global representation. We highlight some of the global results from limited-resource countries here.

Despite the limited evidence for African countries (2 studies [116,121]), our findings were aligned with other African reviews around insufficient evidence of use due to considerable barriers in this region [16] and a lack of "meaningful investment" in this area [173]. We concurred with Nittari et al [174] that several barriers are still present that risk the sustainability of this delivery mode beyond the pandemic.

Studies in South America [46,110,166] provide examples of effective use in outpatient settings (specifically speech and language therapy [46]), as well as reflections on how this new delivery mode in the context of the wider pandemic might have generated an element of mistrust and fear in the clinician-patient relationship [166].

Results from Middle Eastern and Asian countries [41,97,104,105] provide perspectives on the use of telemedicine in orthopedics and hematology while indicating the equivalence of audio and video consultations, an important point to inform telemedicine programs in low-income countries as a way of increasing access to health services.

Positioning Against the Wider and Recent Literature

This review can be positioned in the emerging literature of reviews around telemedicine during the pandemic period [175-183].

A key finding of our study is how patient experience was generally equated with patient satisfaction. Other studies have found these are often used interchangeably [184-186]. However, accepted definitions of patient experience go beyond satisfaction and "focus on individualized care and tailoring of services to meet patients' needs" [187]. This is related to another finding, as the studies are mainly observational with no reference to patient experience frameworks, let alone to some emerging frameworks specific to the digital patient experience [173,188]. Our findings on patient experience are aligned with another recently published review focused on the COVID-19 pandemic, with similar categories and findings [178]. In terms of the clinician-patient relationship, our findings were mixed, but recent reviews found that the relationship was "troubled" telemedicine given both patient and clinician reluctance to use [177]. However, others found the tool useful [106].

de Oliveira Andrade et al [175] focused on the role of legislation in the widespread utilization of telemedicine during the pandemic, finding that regulatory frameworks enabled telemedicine spread in areas related to ethics, reimbursement, data safety, and pandemic-related regulatory relaxation (in the United States, for instance, relaxation of interstate practice was particularly relevant [183]). We found a lack of consensus in terms of sustainability as these financial and regulatory incentives dissipated; however, supportive regulation would be a defining factor in its sustainability. Our evidence seems more nuanced than other recent reviews [177,189]. We are aware specialties like family medicine and general practice seem to have a preference for a particular medium, such as telephone instead of video [190], with more work being undertaken regarding the impacts on quality [191] or equivalence between face-to-face and remote consultations [183].

Regarding the impact on health inequalities, another review identified emerging literature on the opinions of vulnerable populations regarding telemedicine [180]. We saw references [192-195] to an emerging framework for digital health equity [196], which we expect will help address the gaps we identified in the design and evaluation of inclusive digital health care services to address the "self-reinforcing effect of digital and social exclusion" [197] and its impact on health care inequalities in access and experience. Our concerns about implications on health inequality resonate with other similar reviews for telemedicine [179].

XSL•FO
RenderX

Strengths and Limitations

Strengths

We mapped the emerging literature (gray and academic) on telemedicine during the pandemic to a well-established framework using 3 lenses (patient experience, health inequalities, and clinician-patient relationship). We identified this as the only systematic review of its kind. Only 1 other review mapped academic literature on video consultations to the NASSS framework [36] but outside the context of the pandemic [37]. Notably, we effectively reflected the experiences of non–English-speaking countries with literature across the 5 continents, so it is truly global health–oriented, with the added values and perspectives of a diverse, multidisciplinary research team.

When identifying the limited literature at the intersection of health inequalities and telemedicine, we provided a brief taxonomy of potential groups impacted differently. Recognizing the multidimensionality and intersectionality of social exclusion, we show that demographic characteristics such as age, sex, and socioeconomic factors have received some attention, but there is still very limited information and not enough to draw solid conclusions on impacts.

From a methodological standpoint, we provided additional insight on how to integrate effectively other documents and nontraditional voices and experiences into academic research with a reproducible approach.

Limitations of the Sources

The authors sought gray literature directly via Google to ensure a unified source and methodology to identify and capture experiences from non–English-speaking countries. This is a method with limited reproducibility. Researchers interested in gray literature information in English can consult the UK National Grey Literature Collection Funded by Health Education England. We are aware of additional databases with non-English literature that could be used [198]. Emerging literature has covered mostly single-center survey studies with limited sample sizes, reflective of the immediate experiences arising in the context of the pandemic.

Limitations of the Review

We have limited the depth of analysis to accommodate for the extensive scope, in accordance with the JBI guidelines for data extraction [199]. The use of document extracts using the "surrounding keyword" approach needs further development and testing, as we recognize slightly longer extracts are better at conveying enough information to support screening and analysis. We did not consider synchronous text-based communication. The methodology of scoping reviews is still skewed toward evidence from academic publications, which are biased toward researchers from North America and Europe. Although not specified in the current guidance, capping the number of results from traditional academic databases provides a more balanced representation and could reduce duplication while having limited effects on how comprehensive the findings are.

In conclusion, our discussion section has highlighted considerable variation in the emerging literature during the pandemic regarding changes in pathways toward telemedicine. We highlighted the different focus of studies focused on health inequality or outpatient care and the global representation of the studies included (which is a key strength). Of note is the finding equating patient experience with satisfaction, which reflects a potentially limited understanding of sociotechnical views of human-computer interaction and human factors in traditional health service research.

Conclusions

This section is divided into 3 parts comprising a reflection of our positionality as researchers in analyzing these themes as well as further recommendations for research and practice.

Statement of Positionality

Following the method by Pant et al [37], we frame some of our conclusions in the context of our positionality and our aims. We are a diverse group of researchers (with roots in Latin America, the Middle East, Asia, and Africa and with supervisors from the United Kingdom), but our gaze is colonially influenced by our education and current positionality in this country.

By opening our search criteria to non-English documents and gray literature, we succeeded in capturing immediate, emerging experiences at the global level, with 27.6% (37/134) of our documents having no structured abstract and classified as gray literature, with 5 documents from South America and Africa and 13 documents from Asia. The balance, however, is still very skewed toward the United States and English-speaking countries.

Our positionality and knowledge of the UK health system and legislation meant it was difficult to translate these categories to other systems, and we had to modify our parameters and analysis. For instance, although our choice of the UK diversity legislation ("protected characteristics") as a framework for categorizing health inequalities was helpful, the UK definition of "primary care" was not helpful due to its contrast with the definition set out by the WHO (for example in the work by Dimer et al [46] in Brazil, classifying speech and language therapy services as a "primary health care service," which are not available in the United Kingdom in this setting).

Recommendations for Research

From a methodological standpoint, we urge researchers looking to decolonize academic research to test and evaluate further our approach to using gray literature extracts in other languages, particularly in scoping reviews, as it provides that additional level of immediacy with the phenomenon of study. We noticed confusion on the use of patient satisfaction and patient experience, so research should focus on more robust frameworks reviewing the overall patient pathway (away from the evaluation of a remote encounter [184-186]). As expected, our research found mainly descriptive documents, so future research should focus on robust evaluation of clinical outcomes arising from changes in diagnostic and treatment pathways away from face-to-face settings. We recommend future research with a narrower approach to specific population groups and more

XSL•FO
RenderX

focused on access and outcomes (Table S1 in Multimedia Appendix 5). From a sociotechnical research perspective, we recommend future research using modern techniques (natural language processing, data mining, and sentiment analysis) focusing on categories with closer links to human-computer interaction. Following publication, we will publish our data on Open Science Framework [200].

Recommendations for Practice

The impact of telemedicine on patient safety is critical to determine the sustainability of this intervention when contrasting it with the under- or overutilization of resources [201]. If not ruling out the continued use of telemedicine, the authors outlined the importance of further research and refinements to the intervention itself. We found that models such as the Dynamic Sustainability Framework [75] might be useful to support learning and adaptation with care toward the potential disenfranchisement of some patient groups.

Acknowledgments

ChatGPT and Grammarly were used to edit the final draft. Many thanks to Cee MacDonald at Paperaid for the assistance with extract retrieval and text formatting.

This research did not receive any external funding and was undertaken as part of DV's self-funded PhD studies.

Authors' Contributions

These contributions use Contributor Roles Taxonomy (CRediT) guidelines. As this article is also a pedagogic exercise related to various individuals' studies, it was agreed that all voluntary contributors would be set as co-authors regardless of the degree of contribution.
This article contributes to DV's PhD studies at the University of Warwick, and DV contributed to the study conceptualization, data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), project administration (lead), software, validation, visualization, writing of the original manuscript draft (lead), and manuscript review and editing. AS contributed to the data curation (lead), formal analysis (support), investigation (support), methodology (equal), project administration (support), software (support), visualization (support), writing of the original manuscript draft (support), and copyediting (lead). DOM and IL contributed to the data curation (support), investigation (support), and methodology (equal). GH and HZ contributed to the formal analysis (support), visualization (support), and writing of the original manuscript draft (support). TB contributed to the data curation (support), investigation (support), methodology (equal), and manuscript review and editing (support). SAI contributed to the data curation (support), formal analysis (support), and methodology (support). RP contributed to study supervision (lead), validation (lead), and manuscript review and editing (support). RS contributed to study supervision (support), manuscript review and editing (lead), and validation (lead). JD contributed to study supervision (support). AP contributed to validation of data extraction (support) and preparation of the final draft. LJM and KTK contributed to the data curation (support).

Conflicts of Interest

At the time of publication, DV, TB, and HZ worked in the UK National Health Service in the United Kingdom. The research was undertaken independently of DV's, TB's, and HZ's occupations and does not represent the views of their employers. The remaining authors declare no conflict of interest.

Multimedia Appendix 1
Deviations from the protocol.
[DOCX File , 17 KB - ijmr_v14i1e45367_app1.docx]

Multimedia Appendix 2
Example searches.
[DOCX File , 26 KB - ijmr_v14i1e45367_app2.docx]

Multimedia Appendix 3
Abstracts-extracts used in the study.
[XLSX File (Microsoft Excel File), 270 KB - ijmr_v14i1e45367_app3.xlsx]

Multimedia Appendix 4
Data collection template.
[XLSX File (Microsoft Excel File), 460 KB - ijmr_v14i1e45367_app4.xlsx]

XSL•FO
RenderX

Multimedia Appendix 5
Summary characteristics of the included studies.
[DOCX File , 66 KB - ijmr_v14i1e45367_app5.docx]

Multimedia Appendix 6
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) Checklist.
[PDF File (Adobe PDF File), 101 KB - ijmr_v14i1e45367_app6.pdf]

References

1. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020 Mar 11. URL: https://tinyurl.com/tn4rtbse [accessed 2021-02-20]
2. Rational use of personal protective equipment (PPE) for coronavirus disease (COVID-19). World Health Organization. 2020 Mar 19. URL: https://apps.who.int/iris/bitstream/handle/10665/331498/WHO-2019-nCoV-IPCPPE_use-2020.2-eng.pdf [accessed 2021-06-08]
3. Chersich MF, Gray G, Fairlie L, Eichbaum Q, Mayhew S, Allwood B, et al. COVID-19 in Africa: care and protection for frontline healthcare workers. Global Health 2020 May 15;16(1):46 [FREE Full text] [doi: 10.1186/s12992-020-00574-3] [Medline: 32414379]
4. Greenhalgh T, Wherton J, Shaw S, Morrison C. Video consultations for covid-19. BMJ 2020 Mar 12;368:m998. [doi: 10.1136/bmj.m998] [Medline: 32165352]
5. Wherton J, Greenhalgh T. Evaluation of the Near Me video consulting service in Scotland during COVID-19, 2020: Summary Report. Scottish Government. 2021 Mar. URL: https://tinyurl.com/3nmavy8m [accessed 2024-11-06]
6. The NHS: lessons learnt and building for the future. UK Parliament. URL: https://publications.parliament.uk/pa/cm5801/cmselect/cmhealth/320/32009.htm [accessed 2021-06-12]
7. Valdes D, Alqazlan L, Procter R, Dale J. Global evidence on the rapid adoption of telemedicine in primary care during the first year of the COVID-19 pandemic: a scoping review protocol. Research Square. 2021 Mar 16. URL: https://assets-eu.researchsquare.com/files/rs-302879/v1/5668b4f9-5d65-4353-93db-32249840ca3d.pdf [accessed 2024-11-27]
8. Valdes D, Alqazlan L, Procter R, Dale J. Global evidence on the rapid adoption of telemedicine in primary care during the first 2 years of the COVID-19 pandemic: a scoping review protocol. Syst Rev 2022 Jun 19;11(1):124 [FREE Full text] [doi: 10.1186/s13643-022-01934-3] [Medline: 35718770]
9. Rodgers M, Raine G, Thomas S, Harden M, Eastwood A. Informing NHS policy in 'digital-first primary care': a rapid evidence synthesis. Health Serv Deliv Res 2019 Dec;7(41):1-124. [doi: 10.3310/hsdr07410] [Medline: 31869020]
10. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020 Mar 17;4(1):bjgpopen20X101020. [doi: 10.3399/bjgpopen20x101020]
11. Guidance and standard operating procedures: general practice in the context of coronavirus (COVID-19) Version 3.4. Sheffield Local Medical Committee. 2020 Aug 10. URL: https://tinyurl.com/2s362d9z [accessed 2024-11-06]
12. Gausvik C, Jabbarpour Y. COVID-19 Timeline: Centers for Medicare and Medicaid Services (CMS) changes and primary care support were not enough to prevent practice losses. J Am Board Fam Med 2021 Feb;34(Suppl):S7-S9 [FREE Full text] [doi: 10.3122/jabfm.2021.S1.200305] [Medline: 33622808]
13. Southgate G, Yassaee AA, Harmer MJ, Livesey H, Pryde K, Roland D. Use of telemedicine in pediatric services for 4 representative clinical conditions: scoping review. J Med Internet Res 2022 Oct 26;24(10):e38267 [FREE Full text] [doi: 10.2196/38267] [Medline: 36287609]
14. Weiner S. What happens to telemedicine after COVID-19? Association of American Medical Colleges. 2021 Oct 21. URL: https://www.aamc.org/news-insights/what-happens-telemedicine-after-covid-19 [accessed 2022-11-26]
15. Doarn CR. The receding tide: is telemedicine in retreat? Telemed J E Health 2022 Nov 01;28(11):1553-1554. [doi: 10.1089/tmj.2022.29082.editorial] [Medline: 36538750]
16. Dodoo JE, Al-Samarraie H, Alzahrani AI. Telemedicine use in Sub-Saharan Africa: Barriers and policy recommendations for Covid-19 and beyond. Int J Med Inform 2021 Jul;151:104467 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104467] [Medline: 33915421]
17. Pierce W, Schroeder D, Suchecki R. Telehealth in Latin America: progress, challenges, and opportunities in the face of COVID-19. Telehealth Med Today 2021 Jan 28;6(1):1 [FREE Full text] [doi: 10.30953/tmt.v6.238]
18. Aromataris E, Munn Z. JBI manual for evidence synthesis. JBI 2020:1 [FREE Full text] [doi: 10.46658/JBIMES-20-01]
19. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, 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]
20. Pollock D, Peters M, Khalil H, McInerney P, Alexander L, Tricco A, et al. Recommendations for the extraction, analysis, and presentation of results in scoping reviews. JBI Evid Synth 2023 Mar 01;21(3):520-532. [doi: 10.11124/JBIES-22-00123] [Medline: 36081365]

XSL•FO
RenderX

21. Olszewski C, Thomson S, Strauss L, Graham R, Ezzeddine M, Dodenhoff K, et al. Patient experiences with ambulatory telehealth in neurology: results of a mixed-methods study. Neurol Clin Pract 2021 Dec;11(6):484-496 [FREE Full text] [doi: 10.1212/CPJ.0000000000001072] [Medline: 34992956]
22. WHO Global Observatory for eHealth. Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth. World Health Organization. 2010. URL: https://iris.who.int/handle/10665/44497 [accessed 2024-11-06]
23. Adams RJ, Smart P, Huff AS. Shades of grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies. Int J Management Reviews 2016 Apr 19;19(4):432-454. [doi: 10.1111/ijmr.12102]
24. Tyndall J. How low can you go? Toward a hierarchy of Grey Literature. 2008 Presented at: Dreaming 08 – Australian Library and Information Association Biennial Conference; September 2-5, 2008; Alice Springs, NT, Australia URL: https://core.ac.uk/download/pdf/14931482.pdf [doi: 10.14325/mississippi/9781496828118.003.0029]
25. 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]
26. Interim process and methods for developing rapid guidelines on COVID-19. NICE. 2020 Mar 20. URL: https://www.nice.org.uk/process/pmg35 [accessed 2024-11-06]
27. Downes MJ, Mervin MC, Byrnes JM, Scuffham PA. Telemedicine for general practice: a systematic review protocol. Syst Rev 2015 Oct 05;4(1):134 [FREE Full text] [doi: 10.1186/s13643-015-0115-2] [Medline: 26597992]
28. How to compare Ovid MEDLINE and PubMed. Wolters Kluwer Health. 2019 Nov. URL: https://tinyurl.com/442zne52 [accessed 2024-11-06]
29. Assenza G, Lanzone J, Brigo F, Coppola A, Di Gennaro G, Di Lazzaro V, et al. Epilepsy care in the time of COVID-19 pandemic in Italy: risk factors for seizure worsening. Front Neurol 2020;11:737 [FREE Full text] [doi: 10.3389/fneur.2020.00737] [Medline: 32719655]
30. Garousi V, Felderer M, Mäntylä MV. Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. Information and Software Technology 2019 Feb;106:101-121. [doi: 10.1016/j.infsof.2018.09.006]
31. Altingovde IS, Demir E, Can F, Ulusoy Ö. Site-based dynamic pruning for query processing in search engines. 2008 Presented at: SIGIR '08: 31st annual international ACM SIGIR conference on Research and development in information retrieval; July 20-24, 2008; Singapore, Singapore. [doi: 10.1145/1390334.1390543]
32. Szmuda T, Özdemir C, Ali S, Singh A, Syed M, Słoniewski P. Readability of online patient education material for the novel coronavirus disease (COVID-19): a cross-sectional health literacy study. Public Health 2020 Aug;185:21-25 [FREE Full text] [doi: 10.1016/j.puhe.2020.05.041] [Medline: 32516624]
33. Beel J, Gipp B. Google Scholar's Ranking Algorithm: The Impact of Articles' Age (An Empirical Study). 2009 Presented at: Sixth International Conference on Information Technology: New Generations; April 27-29, 2009; Las Vegas, NV. [doi: 10.1109/itng.2009.317]
34. Petticrew M, Roberts H. Systematic Reviews in the Social Sciences: A Practical Guide. Hoboken, NJ: John Wiley & Sons; 2008.
35. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
36. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
37. Pant I, Khosla S, Lama JT, Shanker V, AlKhaldi M, El-Basuoni A, et al. Decolonising global health evaluation: Synthesis from a scoping review. PLOS Glob Public Health 2022 Nov 16;2(11):e0000306 [FREE Full text] [doi: 10.1371/journal.pgph.0000306] [Medline: 36962490]
38. Kumar M, Atwoli L, Burgess RA, Gaddour N, Huang KY, Kola L, et al. What should equity in global health research look like? The Lancet 2022 Jul;400(10347):145-147. [doi: 10.1016/s0140-6736(22)00888-1]
39. van Wyk A. Decolonising research methods and knowledge: implications for stepping into a post-doctoral role. Health & Social Care Workforce. 2022 Aug 02. URL: https://tinyurl.com/34xcnf7c [accessed 2024-11-06]
40. Adams L, Lester S, Hoon E, van der Haak H, Proudman C, Hall C, et al. Patient satisfaction and acceptability with telehealth at specialist medical outpatient clinics during the COVID-19 pandemic in Australia. Intern Med J 2021 Jul;51(7):1028-1037 [FREE Full text] [doi: 10.1111/imj.15205] [Medline: 34213046]
41. Alhajri N, Simsekler MCE, Alfalasi B, Alhashmi M, AlGhatrif M, Balalaa N, et al. Physicians' attitudes toward telemedicine consultations during the COVID-19 pandemic: cross-sectional study. JMIR Med Inform 2021 Jun 01;9(6):e29251 [FREE Full text] [doi: 10.2196/29251] [Medline: 34001497]
42. Alharbi KG, Aldosari MN, Alhassan AM, Alshallal KA, Altamimi AM, Altulaihi BA. Patient satisfaction with virtual clinic during Coronavirus disease (COVID-19) pandemic in primary healthcare, Riyadh, Saudi Arabia. J Family Community Med 2021;28(1):48-54 [FREE Full text] [doi: 10.4103/jfcm.JFCM_353_20] [Medline: 33679189]

XSL•FO
RenderX

43. Aliberti GM, Bhatia R, Desrochers LB, Gilliam EA, Schonberg MA. Perspectives of primary care clinicians in Massachusetts on use of telemedicine with adults aged 65 and older during the COVID-19 pandemic. Prev Med Rep 2022 Apr;26:101729 [FREE Full text] [doi: 10.1016/j.pmedr.2022.101729] [Medline: 35155085]
44. Al-Iede M, Waters K, Aleidi S, Alqutawneh B, Alnawaiseh H, Alshraideh A, et al. Impact of COVID-19 lockdown on children with asthma in Jordan: a parental questionnaire. BMJ Paediatr Open 2021;5(1):e001136 [FREE Full text] [doi: 10.1136/bmjpo-2021-001136] [Medline: 34222681]
45. Altulaihi BA, Alharbi KG, Alhassan AM, Altamimi AM, Al Akeel MA. Physician's perception toward using telemedicine during COVID-19 pandemic in King Abdulaziz Medical City, Riyadh, Saudi Arabia. Cureus 2021 Jul;13(7):e16107 [FREE Full text] [doi: 10.7759/cureus.16107] [Medline: 34350074]
46. Dimer NA, do Canto-Soares N, Dos Santos-Teixeira L, de Goulart BNG. The COVID-19 pandemic and the implementation of telehealth in speech-language and hearing therapy for patients at home: an experience report. Codas 2020;32(3):e20200144 [FREE Full text] [doi: 10.1590/2317-1782/20192020144] [Medline: 32578694]
47. Ball E, Rivas C, Khan R. If virtual gynecology clinics are here to stay, we need to include everyone. AJOG Glob Rep 2022 Feb;2(1):100043 [FREE Full text] [doi: 10.1016/j.xagr.2021.100043] [Medline: 34909705]
48. Barkai G, Gadot M, Amir H, Menashe M, Shvimer-Rothschild L, Zimlichman E. Patient and clinician experience with a rapidly implemented large-scale video consultation program during COVID-19. Int J Qual Health Care 2021 Feb 20;33(1):1 [FREE Full text] [doi: 10.1093/intqhc/mzaa165] [Medline: 33313891]
49. Belenkiy IG. COVID-19 challenge: what has been done and what must be done? Traumatology and Orthopedics of Russia 2020 Jul 06;26(2):15-19. [doi: 10.21823/2311-2905-2020-26-2-15-19]
50. Berg WT, Goldstein M, Melnick AP, Rosenwaks Z. Clinical implications of telemedicine for providers and patients. Fertil Steril 2020 Dec;114(6):1129-1134 [FREE Full text] [doi: 10.1016/j.fertnstert.2020.10.048] [Medline: 33280717]
51. Bergman D, Bethell C, Gombojav N, Hassink S, Stange KC. Physical distancing with social connectedness. Ann Fam Med 2020 May 01;18(3):272-277 [FREE Full text] [doi: 10.1370/afm.2538] [Medline: 32393566]
52. Byrne E, Watkinson S. Patient and clinician satisfaction with video consultations during the COVID-19 pandemic: an opportunity for a new way of working. J Orthod 2021 Mar;48(1):64-73. [doi: 10.1177/1465312520973677] [Medline: 33251951]
53. Chesnel C, Hentzen C, Le Breton F, Turmel N, Tan E, Haddad R, et al. Efficiency and satisfaction with telephone consultation of follow-up patients in neuro-urology: Experience of the COVID-19 pandemic. Neurourol Urodyn 2021 Mar;40(3):929-937 [FREE Full text] [doi: 10.1002/nau.24651] [Medline: 33675263]
54. Chiang C, Halker Singh R, Lalvani N, Shubin Stein K, Henscheid Lorenz D, Lay C, et al. Patient experience of telemedicine for headache care during the COVID-19 pandemic: An American Migraine Foundation survey study. Headache 2021 May;61(5):734-739 [FREE Full text] [doi: 10.1111/head.14110] [Medline: 34021595]
55. La Telemedicina Cara a Cara. Comité de ética asistencial de Atención Primaria de Bizkaia. 2020 Dec. URL: https://osalde.org/wp-content/uploads/2021/01/Telemedicina-aspectos-%C3%A9ticos-CEAAPB.pdf [accessed 2022-10-26]
56. Conway J, Krieger P, Hasanaj L, Sun L, Scharf JM, Odel JG, et al. Telemedicine evaluations in neuro-ophthalmology during the COVID-19 pandemic: patient and physician surveys. J Neuroophthalmol 2021 Sep 01;41(3):356-361. [doi: 10.1097/WNO.0000000000001370] [Medline: 34415269]
57. Curtis M, Duncan R, Jing M, Kim A, Lu VT, Redshaw J, et al. "Not a perfect situation, but..." A single-practice survey of patient experience of phone consultations during COVID-19 Alert Level 4 in New Zealand. N Z Med J 2021 Oct 22;134(1544):35-48. [Medline: 34695091]
58. Day K, Hunter I, Rao V, Russell G, Roskvist R, Moir F, et al. Survey protocol for exploring video and phone use in Aotearoa New Zealand general practice: considerations for future telehealth. BMJ Health Care Inform 2021 Mar;28(1):1 [FREE Full text] [doi: 10.1136/bmjhci-2020-100309] [Medline: 33753326]
59. Dempsey CM, Serino-Cipoletta JM, Marinaccio BD, O'Malley KA, Goldberg NE, Dolan CM, et al. Determining factors that influence parents' perceptions of telehealth provided in a pediatric gastroenterological practice: A quality improvement project. J Pediatr Nurs 2022;62:36-42. [doi: 10.1016/j.pedn.2021.11.023] [Medline: 34894421]
60. Dhahri AA, Iqbal MR, Pardoe H. Agile application of video telemedicine during the COVID-19 pandemic. Cureus 2020 Nov 04;12(11):e11320 [FREE Full text] [doi: 10.7759/cureus.11320] [Medline: 33262918]
61. Dietzen A, Ide W, Pavone L. Telehealth in pediatric rehabilitation medicine: A survey of clinician reported utilization and experience. J Pediatr Rehabil Med 2020;13(3):281-288. [doi: 10.3233/PRM-200762] [Medline: 33252099]
62. Dooley AB, Houssaye NDL, Baum N. Use of telemedicine for sexual medicine patients. Sex Med Rev 2020 Oct;8(4):507-517 [FREE Full text] [doi: 10.1016/j.sxmr.2020.06.001] [Medline: 32739238]
63. Dopelt K, Avni N, Haimov-Sadikov Y, Golan I, Davidovitch N. Telemedicine and eHealth literacy in the era of COVID-19: a cross-sectional study in a peripheral clinic in Israel. Int J Environ Res Public Health 2021 Sep 10;18(18):9556 [FREE Full text] [doi: 10.3390/ijerph18189556] [Medline: 34574480]
64. Board of Directors Meeting - Open (Thursday 27 May 2021). East Kent Hospitals NHS Foundation Trust. 2021. URL: https://www.ekhuft.nhs.uk/about-us/board-of-directors/ [accessed 2024-12-05]

XSL•FO
RenderX

65. Ennis M, Wahl K, Jeong D, Knight K, Renner R, Munro S, et al. The perspective of Canadian health care professionals on abortion service during the COVID-19 pandemic. Fam Pract 2021 Aug 27;38(Suppl 1):i30-i36 [FREE Full text] [doi: 10.1093/fampra/cmab083] [Medline: 34448482]
66. Estudio sobre la consulta de salud virtual y sus beneficios para el sistema sanitario. EsadeGov. 2020 Jul 16. URL: https://tinyurl.com/463ek5dw [accessed 2022-10-26]
67. Esper GJ, Sweeney RL, Winchell E, Duffell JM, Kier SC, Lukens HW, et al. Rapid systemwide implementation of outpatient telehealth in response to the COVID-19 pandemic. J Healthc Manag 2020;65(6):443-452 [FREE Full text] [doi: 10.1097/JHM-D-20-00131] [Medline: 33074968]
68. The organisation of resilient health and social care following the COVID-19 pandemic. European Commission. 2020. URL: https://health.ec.europa.eu/system/files/2020-12/026_health_socialcare_covid19_en_0.pdf [accessed 2022-10-26]
69. Florea M, Lazea C, Gaga R, Sur G, Lotrean L, Puia A, et al. Lights and shadows of the perception of the use of telemedicine by Romanian family doctors during the COVID-19 pandemic. Int J Gen Med 2021;14:1575-1587 [FREE Full text] [doi: 10.2147/IJGM.S309519] [Medline: 33953605]
70. Fluhr JW, Gueguen A, Legoupil D, Brenaut E, Abasq C, Araújo H, et al. Teledermatology in times of COVID-19 confinement: comparing patients' and physicians' satisfaction by the Standardized Brest Teledermatology Questionnaire. Dermatology 2021 Feb 10;237(2):1-6 [FREE Full text] [doi: 10.1159/000514029] [Medline: 33567427]
71. Frank HE, Grumbach NM, Conrad SM, Wheeler J, Wolff J. Mental health services in primary care: Evidence for the feasibility of telehealth during the COVID-19 pandemic. J Affect Disord Rep 2021 Jul;5:100146 [FREE Full text] [doi: 10.1016/j.jadr.2021.100146] [Medline: 33870261]
72. Franzosa E, Gorbenko K, Brody AA, Leff B, Ritchie CS, Kinosian B, et al. "There is something very personal about seeing someone's face": provider perceptions of video visits in home-based primary care during COVID-19. J Appl Gerontol 2021 Nov;40(11):1417-1424 [FREE Full text] [doi: 10.1177/07334648211028393] [Medline: 34210200]
73. Franzosa E, Gorbenko K, Brody AA, Leff B, Ritchie CS, Kinosian B, et al. "At Home, with Care": lessons from New York City home-based primary care practices managing COVID-19. J Am Geriatr Soc 2021 Feb;69(2):300-306 [FREE Full text] [doi: 10.1111/jgs.16952] [Medline: 33179761]
74. Fu L, Lindenmeyer A, Phillimore J, Lessard-Phillips L. Vulnerable migrants' access to healthcare in the early stages of the COVID-19 pandemic in the UK. Public Health 2022 Feb;203:36-42 [FREE Full text] [doi: 10.1016/j.puhe.2021.12.008] [Medline: 35026578]
75. Gold KJ, Laurie AR, Kinney DR, Harmes KM, Serlin DC. Video visits: family physician experiences with uptake during the COVID-19 pandemic. Fam Med 2021 Mar 4;53(3):207-210. [doi: 10.22454/fammed.2021.613099]
76. Gray DP, Sidaway-Lee K, Harding A, Evans P. Reduction in face-to-face GP consultations. Br J Gen Pract 2020 Jul;70(696):328 [FREE Full text] [doi: 10.3399/bjgp20X710849] [Medline: 32586804]
77. Grens H, de Bruin JP, Huppelschoten A, Kremer JAM. Fertility workup with video consultation during the COVID-19 pandemic: pilot quantitative and qualitative study. JMIR Form Res 2022 Feb 07;6(2):e32000. [doi: 10.2196/32000] [Medline: 34936981]
78. Grosman-Dziewiszek P, Wiatrak B, Jęśkowiak I, Szeląg A. Patients' habits and the role of pharmacists and telemedicine as elements of a modern health care system during the COVID-19 pandemic. J Clin Med 2021 Sep 17;10(18):4211 [FREE Full text] [doi: 10.3390/jcm10184211] [Medline: 34575322]
79. Gupta T, Gkiousias V, Bhutta MF. A systematic review of outcomes of remote consultation in ENT. Clin Otolaryngol 2021 Jul;46(4):699-719. [doi: 10.1111/coa.13768] [Medline: 33754458]
80. Hardcastle L, Ogbogu U. Virtual care: Enhancing access or harming care? Healthc Manage Forum 2020 Nov 20;33(6):288-292 [FREE Full text] [doi: 10.1177/0840470420938818] [Medline: 32686506]
81. Hays RD, Skootsky SA. Patient experience with in-person and telehealth visits before and during the COVID-19 pandemic at a large integrated health system in the United States. J Gen Intern Med 2022 Mar;37(4):847-852 [FREE Full text] [doi: 10.1007/s11606-021-07196-4] [Medline: 34982370]
82. Heyck Lee S, Ramondino S, Gallo K, Moist LM. A quantitative and qualitative study on patient and physician perceptions of nephrology telephone consultation during COVID-19. Can J Kidney Health Dis 2022;9:20543581211066720 [FREE Full text] [doi: 10.1177/20543581211066720] [Medline: 35024152]
83. Hofmann G, Hampanda K, Harrison MS, Fasano M, Nacht A, Yeoman M. Virtual prenatal and postpartum care acceptability among maternity care providers. Matern Child Health J 2022 Jul 15;26(7):1401-1408 [FREE Full text] [doi: 10.1007/s10995-022-03412-7] [Medline: 35292887]
84. Holtz BE. Patients perceptions of telemedicine visits before and after the coronavirus disease 2019 pandemic. Telemed J E Health 2021 Jan 01;27(1):107-112 [FREE Full text] [doi: 10.1089/tmj.2020.0168] [Medline: 32614689]
85. Imlach F, McKinlay E, Middleton L, Kennedy J, Pledger M, Russell L, et al. Telehealth consultations in general practice during a pandemic lockdown: survey and interviews on patient experiences and preferences. BMC Fam Pract 2020 Dec 13;21(1):269 [FREE Full text] [doi: 10.1186/s12875-020-01336-1] [Medline: 33308161]
86. Isautier JM, Copp T, Ayre J, Cvejic E, Meyerowitz-Katz G, Batcup C, et al. People's experiences and satisfaction with telehealth during the COVID-19 pandemic in Australia: cross-sectional survey study. J Med Internet Res 2020 Dec 10;22(12):e24531 [FREE Full text] [doi: 10.2196/24531] [Medline: 33156806]

XSL•FO
RenderX

87. Itamura K, Tang DM, Higgins TS, Rimell FL, Illing EA, Ting JY, et al. Comparison of patient satisfaction between virtual visits during the COVID-19 pandemic and in-person visits pre-pandemic. Ann Otol Rhinol Laryngol 2021 Jul;130(7):810-817 [FREE Full text] [doi: 10.1177/0003489420977766] [Medline: 33251849]
88. Javanparast S, Roeger L, Kwok Y, Reed RL. The experience of Australian general practice patients at high risk of poor health outcomes with telehealth during the COVID-19 pandemic: a qualitative study. BMC Fam Pract 2021 Apr 08;22(1):69 [FREE Full text] [doi: 10.1186/s12875-021-01408-w] [Medline: 33832422]
89. Jhaveri D, Emeto TI, Alele FO, Strom A, Benham H. Use of telemedicine for rheumatology practice in Queensland, Australia: experiences before and during the COVID-19 pandemic. Intern Med J 2022 Oct;52(10):1685-1690. [doi: 10.1111/imj.15706] [Medline: 35112769]
90. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of video consultations during the COVID-19 pandemic lockdown: cross-sectional survey among Norwegian general practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
91. Johnson C, Dupuis JB, Goguen P, Grenier G. Changes to telehealth practices in primary care in New Brunswick (Canada): A comparative study pre and during the COVID-19 pandemic. PLoS One 2021;16(11):e0258839 [FREE Full text] [doi: 10.1371/journal.pone.0258839] [Medline: 34813618]
92. Jonnagaddala J, Godinho MA, Liaw S. From telehealth to virtual primary care in Australia? A Rapid scoping review. Int J Med Inform 2021 Jul;151:104470 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104470] [Medline: 34000481]
93. Kalwani NM, Wang KM, Johnson AN, Deb JD, Gold T, Maddukuri AK, et al. Application of the Quadruple Aim to evaluate the operational impact of a telemedicine program. Healthc (Amst) 2021 Dec;9(4):100593 [FREE Full text] [doi: 10.1016/j.hjdsi.2021.100593] [Medline: 34749227]
94. Kamimura A, Panahi S, Meng H, Sundrud J, Lucero M. Patient satisfaction with telehealth and dxperiences during the COVID-19 pandemic among uninsured free clinic patients. J Patient Exp 2021 Jul 26;8:23743735211033107 [FREE Full text] [doi: 10.1177/23743735211033107] [Medline: 34368425]
95. Karacabeyli D, Shea S, Keidar S, Pinkney S, Bepple K, Edwards D, et al. The Live 5-2-1-0 Toolkit for family physicians: Mixed methods evaluation of a resource to facilitate health promotion in a primary care setting. BC Medical Journal 2020;62(6):1 [FREE Full text]
96. Kato-Lin Y, Thelen ST. Telemedicine for acute conditions during COVID-19: a nationwide survey using crowdsourcing. Telemed J E Health 2021 Jul;27(7):714-723. [doi: 10.1089/tmj.2020.0351] [Medline: 33197368]
97. Khan ZA, Zahoor A, Afzal I, Butt U, Siddiqui AM, Khan Z, et al. Evaluation of patient perception and satisfaction toward the use of telemedicine during pandemic of novel coronavirus in Pakistan. Telemed J E Health 2021 Oct;27(10):1174-1179. [doi: 10.1089/tmj.2020.0343] [Medline: 33449871]
98. Kirby DJ, Fried JW, Buchalter DB, Moses MJ, Hurly ET, Cardone DA, et al. Patient and physician satisfaction with telehealth during the COVID-19 pandemic: sports medicine perspective. Telemed J E Health 2021 Oct;27(10):1151-1159. [doi: 10.1089/tmj.2020.0387] [Medline: 33512302]
99. Kletečka-Pulker M, Völkl-Kernstock S, Fassl A, Klager E, Willschke H, Klomfar S, et al. Telehealth in times of COVID-19: spotlight on Austria. Healthcare (Basel) 2021 Mar 04;9(3):1 [FREE Full text] [doi: 10.3390/healthcare9030280] [Medline: 33806673]
100. Kling SMR, Falco-Walter JJ, Saliba-Gustafsson EA, Garvert DW, Brown-Johnson CG, Miller-Kuhlmann R, et al. Patient and clinician perspectives of new and return ambulatory teleneurology visits. Neurol Clin Pract 2021 Dec;11(6):472-483 [FREE Full text] [doi: 10.1212/CPJ.0000000000001065] [Medline: 34992955]
101. Kludacz-Alessandri M, Walczak R, Hawrysz L, Korneta P. The quality of medical care in the conditions of the COVID-19 pandemic, with particular emphasis on the access to primary healthcare and the effectiveness of treatment in Poland. J Clin Med 2021 Aug 09;10(16):3502 [FREE Full text] [doi: 10.3390/jcm10163502] [Medline: 34441798]
102. Kludacz-Alessandri M, Hawrysz L, Korneta P, Gierszewska G, Pomaranik W, Walczak R. The impact of medical teleconsultations on general practitioner-patient communication during COVID- 19: A case study from Poland. PLoS One 2021;16(7):e0254960 [FREE Full text] [doi: 10.1371/journal.pone.0254960] [Medline: 34270587]
103. Krasovsky T, Silberg T, Barak S, Eisenstein E, Erez N, Feldman I, et al. Transition to multidisciplinary pediatric telerehabilitation during the COVID-19 pandemic: strategy development and implementation. Int J Environ Res Public Health 2021 Feb 04;18(4):1 [FREE Full text] [doi: 10.3390/ijerph18041484] [Medline: 33557395]
104. Kumar S, Kumar A, Kumar M, Kumar A, Arora R, Sehrawat R. Feasibility of telemedicine in maintaining follow-up of orthopaedic patients and their satisfaction: A preliminary study. J Clin Orthop Trauma 2020 Oct;11(Suppl 5):S704-S710 [FREE Full text] [doi: 10.1016/j.jcot.2020.07.026] [Medline: 32837105]
105. Kumar P, Aggarwal M, Dhawan R, Dass J, Kumar G, Sharma V, et al. Tele-medicine services in hematological practice during Covid pandemic: its feasibility and difficulties. Indian J Hematol Blood Transfus 2021 Oct;37(4):528-533 [FREE Full text] [doi: 10.1007/s12288-020-01385-7] [Medline: 33250594]
106. Kurotschka PK, Serafini A, Demontis M, Serafini A, Mereu A, Moro MF, et al. General practitioners' experiences during the first phase of the COVID-19 pandemic in Italy: a critical incident technique study. Front Public Health 2021;9:623904 [FREE Full text] [doi: 10.3389/fpubh.2021.623904] [Medline: 33614587]

XSL•FO
RenderX

107. LeBrun DG, Malfer C, Wilson M, Carroll KM, Wang Ms V, Mayman DJ, et al. Telemedicine in an outpatient arthroplasty setting during the COVID-19 pandemic: early lessons from New York City. HSS J 2021 Feb;17(1):25-30 [FREE Full text] [doi: 10.1177/1556331620972659] [Medline: 33967638]
108. Lewis YD, Elran-Barak R, Grundman-Shem Tov R, Zubery E. The abrupt transition from face-to-face to online treatment for eating disorders: a pilot examination of patients' perspectives during the COVID-19 lockdown. J Eat Disord 2021 Mar 05;9(1):31 [FREE Full text] [doi: 10.1186/s40337-021-00383-y] [Medline: 33673876]
109. Li E, Tsopra R, Jimenez G, Serafini A, Gusso G, Lingner H, et al. General practitioners' perceptions of using virtual primary care during the COVID-19 pandemic: An international cross-sectional survey study. PLOS Digit Health 2022 May;1(5):e0000029 [FREE Full text] [doi: 10.1371/journal.pdig.0000029] [Medline: 36812543]
110. Lopez Guzman Y, Barros Castillo SM, Leal Betancourt MI. Condiciones de vulnerabilidad, sociodemográficas y clínicas de la población con COVID-19 incluida en la estrategia de atención primaria en salud. Mosquera, 2020. Universidad del Rosario. 2021 Oct. URL: https://tinyurl.com/2ckxdpnu [accessed 2022-10-26]
111. Maher DP, Hess D, Edwards C, Allen L. Changes in patient satisfaction scores during the early COVID-19 pandemic. J Patient Exp 2021;8:23743735211034610 [FREE Full text] [doi: 10.1177/23743735211034610] [Medline: 34377775]
112. Malliaras P, Merolli M, Williams CM, Caneiro JP, Haines T, Barton C. 'It's not hands-on therapy, so it's very limited': Telehealth use and views among allied health clinicians during the coronavirus pandemic. Musculoskelet Sci Pract 2021 Apr;52:102340 [FREE Full text] [doi: 10.1016/j.msksp.2021.102340] [Medline: 33571900]
113. Manski-Nankervis J, Davidson S, Hiscock H, Hallinan C, Ride J, Lingam V, et al. Primary care consumers' experiences and opinions of a telehealth consultation delivered via video during the COVID-19 pandemic. Aust J Prim Health 2022 Jun;28(3):224-231. [doi: 10.1071/PY21193] [Medline: 35287793]
114. Massaroni V, Delle Donne V, Ciccarelli N, Ciccullo A, Borghetti A, Faliero D, et al. Use of telehealth for HIV care in Italy: Are doctors and patients on the same page? A cross-sectional study. Int J Med Inform 2021 Dec;156:104616. [doi: 10.1016/j.ijmedinf.2021.104616] [Medline: 34695728]
115. Mathew T, Lee PC, Ianno DJ, Benson J. Telehealth and Australian general practice in 2020: A survey exploring patients' perspectives in the Adelaide Hills. Aust J Gen Pract 2021 Oct;50(10):754-759 [FREE Full text] [doi: 10.31128/AJGP-11-20-5727] [Medline: 34590090]
116. Mbunge E, Batani J, Gaobotse G, Muchemwa B. Virtual healthcare services and digital health technologies deployed during coronavirus disease 2019 (COVID-19) pandemic in South Africa: a systematic review. Glob Health J 2022 Jun;6(2):102-113 [FREE Full text] [doi: 10.1016/j.glohj.2022.03.001] [Medline: 35282399]
117. Menhadji P, Patel R, Asimakopoulou K, Quinn B, Khoshkhounejad G, Pasha P, et al. Patients' and dentists' perceptions of tele-dentistry at the time of COVID-19. A questionnaire-based study. J Dent 2021 Oct;113:103782 [FREE Full text] [doi: 10.1016/j.jdent.2021.103782] [Medline: 34400252]
118. Miller MJ, Pak SS, Keller DR, Barnes DE. Evaluation of pragmatic telehealth physical therapy implementation during the COVID-19 pandemic. Phys Ther 2021 Jan 04;101(1):1 [FREE Full text] [doi: 10.1093/ptj/pzaa193] [Medline: 33284318]
119. Mills KT, Peacock E, Chen J, Zimmerman A, He H, Cyprian A, et al. Experiences and beliefs of low-income patients with hypertension in Louisiana and Mississippi during the COVID-19 pandemic. J Am Heart Assoc 2021 Feb 02;10(3):e018510 [FREE Full text] [doi: 10.1161/JAHA.120.018510] [Medline: 33267723]
120. Moore A, Chavez C, Fisher MP. Factors enhancing trust in electronic communication among patients from an internal medicine clinic: qualitative results of the RECEPT study. J Gen Intern Med 2022 Sep 19;37(12):3121-3127 [FREE Full text] [doi: 10.1007/s11606-021-07345-9] [Medline: 35048293]
121. Mostafa PIN, Hegazy AA. Dermatological consultations in the COVID-19 era: is teledermatology the key to social distancing? An Egyptian experience. J Dermatolog Treat 2022 Mar;33(2):910-915. [doi: 10.1080/09546634.2020.1789046] [Medline: 32602763]
122. Gorgojo Martínez JJ, Murillo AZ, Rubio Herrera M, Lesmes IB. TELECONSULTA en Endocrinología y Nutrición en tiempos de la pandemia COVID-19 y más allá. Sociedad Española de Endocrinología y Nutrición. URL: https://tinyurl.com/2jhyx2es [accessed 2024-11-06]
123. Murren-Boezem J, Solo-Josephson P, Zettler-Greeley CM. On-demand, virtual health care during COVID-19: clinician redeployment and telemedicine utilization in a children's health system. Telemed J E Health 2021 Oct;27(10):1111-1116. [doi: 10.1089/tmj.2020.0461] [Medline: 33393879]
124. Mustafa SS, Vadamalai K, Ramsey A. Patient satisfaction with in-person, video, and telephone allergy/immunology evaluations during the COVID-19 pandemic. J Allergy Clin Immunol Pract 2021 May;9(5):1858-1863 [FREE Full text] [doi: 10.1016/j.jaip.2021.01.036] [Medline: 33548516]
125. Taskforce on Telehealth Policy (TTP): Findings and Recommendations. National Committee for Quality Assurance (NCQA). URL: https://tinyurl.com/4j67w4ju [accessed 2022-10-26]
126. Nguyen OT, Watson AK, Motwani K, Warpinski C, McDilda K, Leon C, et al. Patient-level factors associated with utilization of telemedicine services from a free clinic during COVID-19. Telemed J E Health 2022 Apr;28(4):526-534. [doi: 10.1089/tmj.2021.0102] [Medline: 34255572]

XSL•FO
RenderX

127. Olayiwola JN, Magaña C, Harmon A, Nair S, Esposito E, Harsh C, et al. Telehealth as a bright spot of the COVID-19 pandemic: recommendations from the virtual frontlines ("Frontweb"). JMIR Public Health Surveill 2020 Jun 25;6(2):e19045 [FREE Full text] [doi: 10.2196/19045] [Medline: 32479413]
128. O'Neill C, Matias MV, Peixoto VR, O'Neill H, Aguiar P, Castela E. [Teleconsultation in the Portuguese National Health Service during the COVID-19 pandemic: a survey of physicians' views and future implications]. Acta Med Port 2022 Jun 01;35(6):455-467 [FREE Full text] [doi: 10.20344/amp.16724] [Medline: 35300783]
129. La Atención Primaria Rural en España lanza un grito de socorro para su supervivencia. Médicos y Pacientes. URL: https://tinyurl.com/3jna2w96 [accessed 2022-11-05]
130. Orrange S, Patel A, Mack WJ, Cassetta J. Patient satisfaction and trust in telemedicine during the COVID-19 pandemic: retrospective observational study. JMIR Hum Factors 2021 Apr 22;8(2):e28589 [FREE Full text] [doi: 10.2196/28589] [Medline: 33822736]
131. Parker K, Chia M. Patient and clinician satisfaction with video consultations in dentistry - part one: patient satisfaction. Br Dent J 2021 May 27:1-6 [FREE Full text] [doi: 10.1038/s41415-021-3007-y] [Medline: 34045675]
132. Parsons MW, Gardner MM, Sherman JC, Pasquariello K, Grieco JA, Kay CD, et al. Feasibility and acceptance of direct-to-home tele-neuropsychology services during the COVID-19 pandemic. J Int Neuropsychol Soc 2022 Feb;28(2):210-215. [doi: 10.1017/S1355617721000436] [Medline: 33952375]
133. Patel S, Douglas-Moore J. A reflection on an adapted approach from face-to-face to telephone consultations in our Urology Outpatient Department during the COVID-19 pandemic - a pathway for change to future practice? BJU Int 2020 Sep 04;126(3):339-341 [FREE Full text] [doi: 10.1111/bju.15119] [Medline: 32469096]
134. Written evidence submitted by Philips UKI (HSC0978). UK Parliament. 2021 Mar. URL: https://committees.parliament.uk/writtenevidence/25169/pdf/ [accessed 2022-10-18]
135. Rego N, Pereira HS, Crispim J. Perceptions of patients and physicians on teleconsultation at home for diabetes mellitus: survey study. JMIR Hum Factors 2021 Nov 23;8(4):e27873 [FREE Full text] [doi: 10.2196/27873] [Medline: 34817394]
136. Reicher S, Sela T, Toren O. Using telemedicine during the COVID-19 pandemic: attitudes of adult health care consumers in Israel. Front Public Health 2021;9:653553 [FREE Full text] [doi: 10.3389/fpubh.2021.653553] [Medline: 34079784]
137. Reynolds-Wright JJ, Boydell N, Cameron S, Harden J. A qualitative study of abortion care providers' perspectives on telemedicine medical abortion provision in the context of COVID-19. BMJ Sex Reprod Health 2022 Jul;48(3):199-204 [FREE Full text] [doi: 10.1136/bmjsrh-2021-201309] [Medline: 34848554]
138. Riedel L, Uscher-Pines L, Mehrotra A, Busch AB, Barnett ML, Raja P, et al. Use of telemedicine for opioid use disorder treatment - Perceptions and experiences of opioid use disorder clinicians. Drug Alcohol Depend 2021 Nov 01;228:108999 [FREE Full text] [doi: 10.1016/j.drugalcdep.2021.108999] [Medline: 34517225]
139. Rosler G. Pediatric telehealth experiences: myths and truths about video visits from a parent. J Patient Exp 2020 Dec;7(6):836-838 [FREE Full text] [doi: 10.1177/2374373520932724] [Medline: 33457506]
140. General practice in the post Covid world: challenges and opportunities for general practice. Royal College of General Practitioners. URL: https://tinyurl.com/4kkaxtxy [accessed 2024-11-06]
141. Focus: artificial intelligence and technology in family medicine. Family Doctor. 2021. URL: https://www.nysafp.org/wp-content/uploads/2024/05/Winter-2021.pdf [accessed 2024-11-06]
142. Saiyed S, Nguyen A, Singh R. Physician perspective and key satisfaction indicators with rapid telehealth adoption during the coronavirus disease 2019 pandemic. Telemed J E Health 2021 Nov;27(11):1225-1234. [doi: 10.1089/tmj.2020.0492] [Medline: 33513045]
143. Scherer J, Back DA, Thienemann F, Kaufmann E, Neuhaus V, Willy C, et al. The effect of Covid-19 on the willingness to use video consultations among orthopedic and trauma outpatients: a multi-center survey in 1400 outpatients. Eur J Trauma Emerg Surg 2022 Jun;48(3):2199-2206 [FREE Full text] [doi: 10.1007/s00068-021-01774-1] [Medline: 34427693]
144. Schweiberger K, Hoberman A, Iagnemma J, Schoemer P, Squire J, Taormina J, et al. Practice-level variation in telemedicine use in a pediatric primary care network during the COVID-19 pandemic: retrospective analysis and survey study. J Med Internet Res 2020 Dec 18;22(12):e24345 [FREE Full text] [doi: 10.2196/24345] [Medline: 33290244]
145. Shachak A, Alkureishi MA. Virtual care: a 'Zoombie' apocalypse? J Am Med Inform Assoc 2020 Nov 01;27(11):1813-1815 [FREE Full text] [doi: 10.1093/jamia/ocaa185] [Medline: 32940711]
146. Sharma AE, Khoong EC, Sierra M, Rivadeneira NA, Nijagal MA, Su G, et al. System-level factors associated with telephone and video visit use: survey of safety-net clinicians during the early phase of the COVID-19 pandemic. JMIR Form Res 2022 Mar 10;6(3):e34088 [FREE Full text] [doi: 10.2196/34088] [Medline: 35148271]
147. Skelly JR, O'Connor T. Guidelines for the use of the Attend Anywhere Platform for Telecommunications within the pain service. Ir Med J 2021 Aug 19;114(7):403. [Medline: 34520346]
148. Srinivasan M, Asch S, Vilendrer S, Thomas SC, Bajra R, Barman L, et al. Qualitative assessment of rapid system transformation to primary care video visits at an academic medical center. Annals of Internal Medicine 2020 Oct 06;173(7):527-535. [doi: 10.7326/m20-1814]
149. Staffieri SE, Mathew AA, Sheth SJ, Ruddle JB, Elder JE. Parent satisfaction and acceptability of telehealth consultations in pediatric ophthalmology: initial experience during the COVID-19 pandemic. J AAPOS 2021 Apr;25(2):104-107. [doi: 10.1016/j.jaapos.2020.12.003] [Medline: 33689911]

XSL•FO
RenderX

150. Sugarman DE, Horvitz LE, Greenfield SF, Busch AB. Clinicians' perceptions of rapid scale-up of telehealth services in outpatient mental health treatment. Telemed J E Health 2021 Dec;27(12):1399-1408 [FREE Full text] [doi: 10.1089/tmj.2020.0481] [Medline: 33600272]
151. Hamui-Sutton L, Halabe-Cherem J. Videoconsultas: alcances y limitaciones. Medicina Interna de México 2021;37(2):261-266 [FREE Full text]
152. Swaminathan R, Mughal Z, Phillips D. Telephone Consultation in Otorhinolaryngology During the Coronavirus Disease 2019 Pandemic: a Cross-sectional Analysis of Effectiveness and Satisfaction for Patients and Clinicians. SN Compr Clin Med 2022;4(1):36 [FREE Full text] [doi: 10.1007/s42399-022-01119-y] [Medline: 35036851]
153. Tai-Seale M, Rosen R, Ruo B, Hogarth M, Longhurst CA, Lander L, et al. Implementation of patient engagement tools in electronic health records to enhance patient-centered communication: protocol for feasibility evaluation and preliminary results. JMIR Res Protoc 2021 Aug 26;10(8):e30431 [FREE Full text] [doi: 10.2196/30431] [Medline: 34435960]
154. Taylor P, Berg C, Thompson J, Dean K, Yuan T, Nallamshetty S, et al. Effective access to care in a crisis period: hypertension control during the COVID-19 pandemic by telemedicine. Mayo Clin Proc Innov Qual Outcomes 2022 Feb;6(1):19-26 [FREE Full text] [doi: 10.1016/j.mayocpiqo.2021.11.006] [Medline: 34805763]
155. Tenforde AS, Iaccarino MA, Borgstrom H, Hefner JE, Silver J, Ahmed M, et al. Telemedicine during COVID-19 for outpatient sports and musculoskeletal medicine physicians. PM R 2020 Sep;12(9):926-932 [FREE Full text] [doi: 10.1002/pmrj.12422] [Medline: 32424977]
156. Teng T, Sareidaki DE, Chemaly N, Bar C, Coste-Zeitoun D, Kuchenbuch M, et al. Physician and patient satisfaction with the switch to remote outpatient encounters in epilepsy clinics during the Covid-19 pandemic. Seizure 2021 Oct;91:60-65 [FREE Full text] [doi: 10.1016/j.seizure.2021.05.013] [Medline: 34098318]
157. Thiele T, Beider S, Kühl H, Mielke G, Holz A, Hirsch S, et al. [Care of rheumatology patients during the lockdown in early 2020 : Telemedicine, delegation, patient satisfaction and vaccination behavior]. Z Rheumatol 2022 Mar;81(2):157-163 [FREE Full text] [doi: 10.1007/s00393-021-01005-3] [Medline: 33974131]
158. Tornero-Molina J, Sánchez-Alonso F, Fernández-Prada M, Bris-Ochaita M, Sifuentes-Giraldo A, Vidal-Fuentes J. Tele-rheumatology during the COVID-19 pandemic. Reumatol Clin (Engl Ed) 2020 Oct 29;18(3):157-163 [FREE Full text] [doi: 10.1016/j.reuma.2020.10.004] [Medline: 33214110]
159. Tozour JN, Bandremer S, Patberg E, Zavala J, Akerman M, Chavez M, et al. Application of telemedicine video visits in a maternal-fetal medicine practice at the epicenter of the COVID-19 pandemic. Am J Obstet Gynecol MFM 2021 Nov;3(6):100469 [FREE Full text] [doi: 10.1016/j.ajogmf.2021.100469] [Medline: 34450341]
160. Triantafillou V, Layfield E, Prasad A, Deng J, Shanti RM, Newman JG, et al. Patient perceptions of head and neck ambulatory telemedicine visits: a qualitative study. Otolaryngol Head Neck Surg 2021 May 14;164(5):923-931. [doi: 10.1177/0194599820943523] [Medline: 32662749]
161. Troncone A, Cascella C, Chianese A, Zanfardino A, Casaburo F, Piscopo A, et al. Doctor-patient relationship in synchronous/real-time video-consultations and in-person visits: an investigation of the perceptions of young people with type 1 diabetes and their parents during the COVID-19 pandemic. Int J Behav Med 2022 Oct;29(5):638-647 [FREE Full text] [doi: 10.1007/s12529-021-10047-5] [Medline: 35076861]
162. Tulupova E, Dobiášová K, Kopsa Těšinová J. Patient experience with telemedicine in primary care during the COVID pandemic: a qualitative study. European Journal of Public Health 2021;31(Supplement_3):ckab165.162. [doi: 10.1093/eurpub/ckab165.162]
163. Utley LM, Manchala GS, Phillips MJ, Doshi CP, Szatalowicz VL, Boozer JR. Bridging the telemedicine gap among seniors during the COVID-19 pandemic. J Patient Exp 2021;8:23743735211014036 [FREE Full text] [doi: 10.1177/23743735211014036] [Medline: 34179442]
164. Verduzco-Gutierrez M, Bean AC, Tenforde AS, Tapia RN, Silver JK. How to conduct an outpatient telemedicine rehabilitation or prehabilitation visit. PM R 2020 Jul 07;12(7):714-720. [doi: 10.1002/pmrj.12380] [Medline: 32297458]
165. Verma P, Kerrison R. Patients' and physicians' experiences with remote consultations in primary care during the COVID-19 pandemic: a multi-method rapid review of the literature. BJGP Open 2022 Jun;6(2):1 [FREE Full text] [doi: 10.3399/BJGPO.2021.0192] [Medline: 35031558]
166. Villalobos J, Calvanese N. [The impact of COVID-19 pandemic on doctor-patient relationship]. Rev Med Chil 2021 Jul;149(7):1070-1074 [FREE Full text] [doi: 10.4067/s0034-98872021000701070] [Medline: 34751309]
167. Volandes AE, Zupanc SN, Paasche-Orlow MK, Lakin JR, Chang Y, Burns EA, et al. Association of an advance care planning video and communication intervention with documentation of advance care planning among older adults: a nonrandomized controlled trial. JAMA Netw Open 2022 Feb 01;5(2):e220354 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.0354] [Medline: 35201306]
168. Vongsachang H, Lagstein O, Boland MV, Repka MX, Kraus CL, Collins ME. Telemedicine utilization by pediatric ophthalmologists during the COVID-19 pandemic. J AAPOS 2021 Oct;25(5):293-295.e1 [FREE Full text] [doi: 10.1016/j.jaapos.2021.05.018] [Medline: 34600105]
169. Webber EC, McMillen BD, Willis DR. Health care disparities and access to video visits before and after the COVID-19 pandemic: findings from a patient survey in primary care. Telemed J E Health 2022 May;28(5):712-719. [doi: 10.1089/tmj.2021.0126] [Medline: 34449270]

XSL•FO
RenderX

170. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 2011 Feb 20;30(4):377-399. [doi: 10.1002/sim.4067] [Medline: 21225900]
171. Wilson G, Windner Z, Bidwell S, Currie O, Dowell A, Halim AA, et al. 'Here to stay': changes to prescribing medication in general practice during the COVID-19 pandemic in New Zealand. J. Prim. Health Care 2021 Aug 13;13(3):222-230. [doi: 10.1071/hc21035]
172. Hoda R. Socio-technical grounded theory for software engineering. IIEEE Trans. Software Eng 2022 Oct 1;48(10):3808-3832. [doi: 10.1109/tse.2021.3106280]
173. Chitungo I, Mhango M, Dzobo M, Denhere K, Chimene M, Musuka G, et al. Towards virtual doctor consultations: A call for the scale-up of telemedicine in sub-Saharan Africa during COVID-19 lockdowns and beyond. Smart Health (Amst) 2021 Jul;21:100207 [FREE Full text] [doi: 10.1016/j.smhl.2021.100207] [Medline: 36570120]
174. Nittari G, Khuman R, Baldoni S, Pallotta G, Battineni G, Sirignano A, et al. Telemedicine practice: review of the current ethical and legal challenges. Telemed J E Health 2020 Dec;26(12):1427-1437 [FREE Full text] [doi: 10.1089/tmj.2019.0158] [Medline: 32049608]
175. de Oliveira Andrade A, Soares AB, de Andrade Palis A, Cabral AM, Barreto CGL, de Souza DB, et al. On the use of telemedicine in the context of COVID-19: legal aspects and a systematic review of technology. Res. Biomed. Eng 2021 Mar 13;38(1):209-227. [doi: 10.1007/s42600-021-00133-8]
176. Ftouni R, AlJardali B, Hamdanieh M, Ftouni L, Salem N. Challenges of telemedicine during the COVID-19 pandemic: a systematic review. BMC Med Inform Decis Mak 2022 Aug 03;22(1):207 [FREE Full text] [doi: 10.1186/s12911-022-01952-0] [Medline: 35922817]
177. Pogorzelska K, Chlabicz S. Patient satisfaction with telemedicine during the COVID-19 pandemic-a systematic review. Int J Environ Res Public Health 2022 May 17;19(10):6113 [FREE Full text] [doi: 10.3390/ijerph19106113] [Medline: 35627650]
178. Doraiswamy S, Abraham A, Mamtani R, Cheema S. Use of telehealth during the COVID-19 pandemic: scoping review. J Med Internet Res 2020 Dec 01;22(12):e24087 [FREE Full text] [doi: 10.2196/24087] [Medline: 33147166]
179. Silva A, Sindico S, Carneiro A, Henrique S, Fernandes A, Gomes J, et al. COVID-19 remote consultation services and population in health inequity-concentrating territories: a scoping review. Telemed J E Health 2021 Aug;27(8):881-897 [FREE Full text] [doi: 10.1089/tmj.2021.0145] [Medline: 34232749]
180. Boland J, Banks S, Krabbe R, Lawrence S, Murray T, Henning T, et al. A COVID-19-era rapid review: using Zoom and Skype for qualitative group research. Public Health Res Pract 2022 Jun 15;32(2):1 [FREE Full text] [doi: 10.17061/phrp31232112] [Medline: 35702751]
181. Andrews E, Berghofer K, Long J, Prescott A, Caboral-Stevens M. Satisfaction with the use of telehealth during COVID-19: An integrative review. Int J Nurs Stud Adv 2020 Nov;2:100008 [FREE Full text] [doi: 10.1016/j.ijnsa.2020.100008] [Medline: 33083791]
182. Beattie M, Lauder W, Atherton I, Murphy DJ. Instruments to measure patient experience of health care quality in hospitals: a systematic review protocol. Syst Rev 2014 Jan 04;3(1):4 [FREE Full text] [doi: 10.1186/2046-4053-3-4] [Medline: 24387141]
183. Shaver J. The state of telehealth before and after the COVID-19 pandemic. Prim Care 2022 Dec;49(4):517-530 [FREE Full text] [doi: 10.1016/j.pop.2022.04.002] [Medline: 36357058]
184. Williams B, Coyle J, Healy D. The meaning of patient satisfaction: an explanation of high reported levels. Soc Sci Med 1998 Nov;47(9):1351-1359. [doi: 10.1016/s0277-9536(98)00213-5] [Medline: 9783878]
185. Williams S, Weinman J, Dale J. Doctor-patient communication and patient satisfaction: a review. Fam Pract 1998 Oct;15(5):480-492. [doi: 10.1093/fampra/15.5.480] [Medline: 9848436]
186. Wolf JA, Niederhauser V, Marshburn D, LaVela SL. Defining patient experience. Patient Experience Journal 2014;1(1):7-19. [doi: 10.35680/2372-0247.1004]
187. Garcia R, Adelakun O. A conceptual framework and pilot study for examining telemedicine satisfaction research. J Med Syst 2019 Jan 25;43(3):51. [doi: 10.1007/s10916-019-1161-4] [Medline: 30684065]
188. Wang T, Giunti G, Melles M, Goossens R. Digital patient experience: umbrella systematic review. J Med Internet Res 2022 Aug 04;24(8):e37952 [FREE Full text] [doi: 10.2196/37952] [Medline: 35925651]
189. Greenhalgh T, Ladds E, Hughes G, Moore L, Wherton J, Shaw SE, et al. Why do GPs rarely do video consultations? qualitative study in UK general practice. Br J Gen Pract 2022 Feb 16;72(718):e351-e360. [doi: 10.3399/bjgp.2021.0658]
190. Mohtar LN, Badr NG. Digital Bridge or Tradeoff: Telehealth Adoption and Healthcare Service Quality. A Scoping Review. In: Za S, Winter R, Lazazzara A, editors. Sustainable Digital Transformation. Lecture Notes in Information Systems and Organisation, vol 59. Cham, Switzerland: Springer International Publishing; 2023:253-268.
191. Lauffenburger JC, Khatib R, Siddiqi A, Albert MA, Keller PA, Samal L, et al. Reducing ethnic and racial disparities by improving undertreatment, control, and engagement in blood pressure management with health information technology (REDUCE-BP) hybrid effectiveness-implementation pragmatic trial: Rationale and design. Am Heart J 2023 Jan;255:12-21 [FREE Full text] [doi: 10.1016/j.ahj.2022.10.003] [Medline: 36220355]

XSL•FO
RenderX

192. Jaworski BK, Webb Hooper M, Aklin WM, Jean-Francois B, Elwood WN, Belis D, et al. Advancing digital health equity: Directions for behavioral and social science research. Transl Behav Med 2023 Apr 03;13(3):132-139 [FREE Full text] [doi: 10.1093/tbm/ibac088] [Medline: 36318232]
193. Kelly M, Fullen B, Martin D, Bradley C, O'Mahony B, McVeigh J. Design and development of an eHealth intervention to support self-management in people with musculoskeletal disorders - 'eHealth: It's TIME': a study protocol. HRB Open Res 2022 Nov 14;5:73. [doi: 10.12688/hrbopenres.13611.1]
194. Caillaud C, Lazarevic N, Pizzuti C, Rosic G, Boehm C, Williams K. Digital Health as a silver lining: exploring pregnancy digital health usage amidst the COVID-19 pandemic. Digitial Health Week 2023 Dec 19;2(1):1 [FREE Full text] [doi: 10.21203/rs.3.rs-2115070/v1]
195. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022 Aug 18;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
196. Ragnedda M, Ruiu ML, Addeo F. The self-reinforcing effect of digital and social exclusion: The inequality loop. Telematics and Informatics 2022 Aug;72:101852. [doi: 10.1016/j.tele.2022.101852]
197. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021 Jan 26;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
198. Walpole SC. Including papers in languages other than English in systematic reviews: important, feasible, yet often omitted. J Clin Epidemiol 2019 Jul;111:127-134. [doi: 10.1016/j.jclinepi.2019.03.004] [Medline: 30951800]
199. Peters M, Marnie C, Tricco A, Pollock D, Munn Z, Alexander L. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020:18-26. [doi: 10.11124/jbies-20-00167]
200. Valdes D. Global evidence on the rapid adoption of telemedicine in primary care in the COVID-19 pandemic. OSF. 2023 Jan 02. URL: https://osf.io/4z5ut/ [accessed 2024-11-06]
201. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 02;8(1):117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]

Abbreviations

JBI: Joanna Briggs Institute
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
NHS: National Health Service
NICE: National Institute for Health and Care Excellence
PCC: Population, Concept, Context
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 27.12.22; peer-reviewed by S Vanderhout, A Frank, E Hadley; comments to author 06.04.23; revised version received 29.07.23; accepted 25.07.24; published 28.02.25.

Please cite as:
Valdes D, Shanker A, Hijazi G, Mensah DO, Bockarie T, Lazar I, Ibrahim SA, Zolfagharinia H, Procter R, Spencer R, Dale J, Paule A, Medlin LJ, Tharuvara Kallottil K
Global Evidence on the Sustainability of Telemedicine in Outpatient and Primary Care During the First 2 Years of the COVID-19 Pandemic: Scoping Review Using the Nonadoption, Abandonment, Scale-Up, Spread, and Sustainability (NASSS) Framework
Interact J Med Res 2025;14:e45367
URL: https://www.i-jmr.org/2025/1/e45367
doi:10.2196/45367
PMID:40053716

©Daniela Valdes, Ankit Shanker, Ghofran Hijazi, Daniel Opoku Mensah, Tahir Bockarie, Ioana Lazar, Siti Aishah Ibrahim, Hamid Zolfagharinia, Rob Procter, Rachel Spencer, Jeremy Dale, Armina Paule, Liam Jonathon Medlin, Keerthana Tharuvara Kallottil. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 28.02.2025. 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

Review

Therapeutic Guidelines for the Self-Management of Major Depressive Disorder: Scoping Review

Priscila de Campos Tibúrcio[1], MS; Priscila Maria Marcheti[2], PhD; Daniela Miori Pascon[3], PhD; Marco Antônio Montebello Junior[4], MBA; Maria Alzete de Lima[5], PhD; Carla Sílvia Fernandes[6], PhD; Célia Samarina Vilaça de Brito Santos[6], PhD; Maria do Perpétuo Socorro de Sousa Nóbrega[1], PhD

[1]School of Nursing, University of São Paulo, São Paulo, Brazil

[2]School of Nursing, Federal University of Mato Grosso do Sul, Campo Grande, Brazil

[3]School of Nursing, Pontifical Catholic University of São Paulo, Sorocaba, Brazil

[4]Sorocaba Engineering Department, FACENS University Center, Sorocaba, Brazil

[5]School of Nursing, Federal University of Rio Grande do Norte, Rio Grande do Norte, Natal, Brazil

[6]Nursing Department, Porto School of Nursing, Porto, Portugal

Corresponding Author:
Priscila de Campos Tibúrcio, MS
School of Nursing
University of São Paulo
Av. Dr. Enéas de Carvalho Aguiar, 419
São Paulo, 05403-000
Brazil
Phone: 55 19997889041
Email: priscila.campostib13@gmail.com

Abstract

Background: Major depressive disorder contributes to the global burden of mental illness. Therapeutic guidelines promote treatment self-management and support caregivers and family members in this process.

Objective: We aimed to identify therapeutic guidelines for the symptoms of major depressive disorder.

Methods: This scoping review followed the assumptions established by the Joanna Briggs Institute and the PRISMA-ScR (Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews) protocol, carried out in 12 databases (LILACS, PubMed, SciELO, Scopus, Web of Science, b-on, BDENF, AgeLine, Cochrane, BVS, IBECS, and CINAHL) and 5 secondary gray literature sources (Google Scholar, Global ETD Search, EBSCO Open Dissertations, CAPES Catalog of Theses and Dissertations, and the Digital Library of Theses and Dissertations of the University of Sao Paulo). The eligibility criteria were based on the population, concept, and context framework: people diagnosed with major depressive disorder aged >18 years (population), therapeutic guidelines for self-management of major depressive disorder symptoms (concept), and symptoms of major depressive disorder (context). Data collection was carried out from March to July 2022 and updated in June 2024. The included studies were experimental, quasi-experimental, analytical observational, descriptive observational, qualitative, or quantitative studies; systematic reviews and meta-analyses; and scoping and literature reviews published in full without time restrictions in English, Spanish, or Portuguese. All the information, as well as the studies captured, was stored in a Microsoft Excel spreadsheet using Rayyan and the *JBI Manual for Evidence Synthesis*. The titles, abstracts, and full texts were carefully read and classified, extracting the results. After review by 2 independent researchers, 62 studies were selected. The results are presented descriptively, including characterization of the studies and mapping and categorization of groups and subgroups of therapeutic guidelines for self-management of major depressive disorder.

Results: In total, 62 studies published between 2011 and 2023 were included, where 44 (71%) came from indexed data sources and 18 (29%) were gray literature indexed on Google Scholar (13/62, 21%), doctoral theses (3/62, 5%), and master's dissertations (2/62, 3%). Among the therapeutic guidelines identified, mapped, and categorized, 7 major groups were identified for self-management: psychotherapy (32/62, 52%), adoption of healthy habits (25/62, 40%), integrative and complementary practices (17/62, 27%), relaxation techniques (9/62, 14%), consultation with a health professional (14/62, 22%), pharmacological therapy (9/62, 14%), and leisure or pleasurable activities (4/62, 6%).

XSL•FO
RenderX

Conclusions: It was possible to identify therapeutic guidelines to promote self-management of major depressive disorder in the adult population. Therapeutic guidance is an important resource for patients, their families, and the community, making patients the protagonists of their own health. For health professionals, therapeutic guidelines become tools that help develop skills and competencies for care among patients, thus ensuring their ability to self-manage major depressive disorder.

(Interact J Med Res 2025;14:e63959) doi:10.2196/63959

KEYWORDS

major depressive disorder; nursing; revision; self-management; symptoms; PRISMA

Introduction

Background

Major depressive disorder is one of the most prevalent mental disorders in the world, as well as one of the clinical conditions that contributes most to the global burden of mental illnesses [1]. The number of people presenting with symptoms suggestive of this disorder has shown a significant increase at an alarming and worrying rate, and it is estimated that it will be the world's first major disabling public health problem by 2030 [2].

It is a classic mental disorder that involves evident alterations in affect, cognition, and neurovegetative functions, affecting approximately 300 million people worldwide [2]. It is estimated that 3.8% of cases lead to functional incapacity and damage to physical and mental health, as well as professional losses and considerable morbidity and mortality due to suicide or association with other illnesses [3].

In addition to these impacts, it is difficult to diagnose major depressive disorder accurately and quickly as its classification and assessment are based on clinical findings and the patient's history [2]. Therapies for the treatment of major depressive disorder, such as psychobiosocial therapies, psychotherapy, and pharmacotherapy, have advanced and are used according to the severity of the symptoms and potential adverse events [4]. Considering that mental health care has undergone important transformations, such as the creation of an out-of-hospital network made up of substitute services that propose the rescue of the singularity or subjectivity of the person in psychological distress, respecting their integrality and existence [5], promoting their autonomy in the management of the illness or treatment means encouraging empowerment, the ability to make choices, and taking responsibility for themselves as a citizen and social being [6].

This knowledge and skills are important tools to prevent complications, control symptoms, self-manage treatment, and even avoid recurrences and hospitalizations [7]. Due to the complexity of the disease, some adaptive actions can be taken, including controlling destabilizing factors, preserving self-confidence, re-establishing important relationships, attempting to regain functioning, managing symptoms, negotiating the care environment, and maintaining satisfying relationships [7,8].

Society has been moving toward the construction and adoption of technological innovations using smartphones and computer applications. These resources can be of great value in the process of promoting people's skills in the mental health field [8]. In this sense, based on the results of this scoping review, a systematic set of therapeutic guidelines was developed based on clinical practice and scientific evidence. This set of therapeutic guidelines is capable of supporting a computational application, generating knowledge in the area of mental health, creating a connection with the patient, and supporting self-management of one's health remotely [8,9].

These therapeutic guidelines can be adopted with the aim of promoting and supporting self-management in the daily treatment of the patient and supporting the caregiver or family members in performing techniques safely and appropriately, completing therapeutic sessions, maintaining functionality, or preventing possible complications of the disorder [9]. It should be emphasized that, although it is important to encourage self-management by the individual and the family, it is also essential that the care provided by health professionals, including nurses, assists in the promotion and reorganization of the self, breaks and overcomes relationships of dependence, and levels the acquisition or development of knowledge and skills for self-management [7,9].

Given the extent of the symptomatology and the particularities of major depressive disorder, a scoping review allows for an exhaustive analysis of the studies available so that the patient has access to information that will allow them to develop attitudes, knowledge, and skills to self-manage based on the best scientific evidence.

Objectives

Similarly, a set of therapeutic guidelines will constitute an instrument for the clinical practice of nurses and of the multidisciplinary team as the deepening of specific skills will allow for the provision of individualized care that is more suited to the needs of the person with major depressive disorder. Therefore, this scoping review was carried out with the aim of identifying therapeutic guidelines in scientific production for the self-management of major depressive disorder symptoms.

Methods

Overview

A scoping review is used to map key concepts, examine existing evidence before conducting a systematic review, and clarify and define conceptual boundaries. This scoping review was prepared in accordance with the methodological proposal of the Joanna Briggs Institute (JBI) in 2020, supported by the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) [10] and PRISMA extension for Scoping Reviews (PRISMA-ScR) [11] guidelines based on the following steps: formulation of the research question, identification of

XSL•FO
RenderX

relevant studies, selection of studies, extraction and analysis of data, and synthesis and construction of the report. The PRISMA-ScR checklist is available in Multimedia Appendix 1. This review protocol was registered in the Open Science Framework platform [12] on November 11, 2023, and was modified on February 19, 2025. There was no deviation from the protocol registered on the Open Science Framework platform.

Research Question

To formulate the guiding research question and guide data collection, the population, concept, and context (PCC) framework was used. The PCC strategy was adopted to drive the research question of this scoping review. In this study, *population* refers to people diagnosed with major depressive disorder aged >18 years, *concept* refers to therapeutic guidelines for self-management of major depressive disorder symptoms, and *context* refers to major depressive disorder symptoms (eg, changes in weight, appetite, and sleep; feelings of guilt; irritability or bad mood; anhedonia; fatigue; and low self-esteem).

Therefore, the following guiding question was defined: "What therapeutic guidelines support self-management of changes in weight, appetite, sleep, feelings of guilt, irritability or bad mood, anhedonia, fatigue and low self-esteem?"

Study Design and Eligibility Criteria

The data search was carried out in 12 information sources (LILACS, PubMed, SciELO, Scopus, Web of Science, b-on, BDENF, AgeLine, Cochrane, BVS, IBECS, and CINAHL) and 5 secondary information sources of gray literature (Google Scholar, Global ETD Search, EBSCO Open Dissertations, CAPES Catalog of Theses and Dissertations, and the Digital Library of Theses and Dissertations of the University of Sao Paulo).

The search strategy was developed using the controlled and noncontrolled descriptors obtained in the initial search plus the Boolean operators "OR," "NOT," and "AND," as well as keywords found in the Health Sciences Descriptors and Medical Subject Headings of the US National Library of Medicine combined with each other according to each database. The strategies used and how they were combined are available in Multimedia Appendix 2.

The search followed five distinct phases according to the JBI methodology, and a team of 5 researchers was assembled for this scoping review: (1) initial search in the selected databases to identify articles on the topic and, from there, select words and indexing terms contained in these publications to develop the full search strategy; (2) use of the keywords and indexing terms identified to search all the databases included; (3) definition of the study design and eligibility criteria, including the determination of the type of study, target population, interventions and outcomes analyzed. In addition, inclusion and exclusion criteria were established for the selection of studies; (4) identification and selection of studies, including the systematic search in databases and sources of gray literature, application of filters and search strategies to ensure the identification of relevant studies and screening of relevant articles (reading of titles, abstracts and, when necessary, the full text); and (5) data extraction and analysis, including the collection of information from the selected studies, the assessment of methodological quality and the synthesis of findings.

The eligibility criteria are described in Textbox 1, and the studies found were refined based on the PCC acronym, type of study, year of publication, language, and results. The search was carried out from March 2022 to July 2022 and updated in June 2024.

XSL•FO
RenderX

Textbox 1. Eligibility criteria.

Inclusion criteria

- Population: people diagnosed with major depressive disorder aged >18 years
- Concept: therapeutic guidelines for the self-management of symptoms of major depressive disorder
- Context: symptoms of major depressive disorder, such as changes in weight, appetite, and sleep; feelings of guilt; irritability or bad mood; anhedonia; fatigue; and low self-esteem
- Type of study: experimental and quasi-experimental studies (randomized and nonrandomized controlled trials), analytical observational studies (prospective and retrospective cohort studies, case-control studies, and analytical cross-sectional studies), descriptive observational studies (case series, individual case reports, and descriptive cross-sectional studies), qualitative and quantitative studies, systematic reviews, meta-analyses, scoping reviews, and literature reviews; due to the scarcity of publications found in the databases, it was also decided to include gray literature (course completion studies, dissertations, and theses)
- Year: there were no restrictions on the search period as the first search strategies developed yielded few studies on the subject
- Language: English, Spanish, and Portuguese
- Results: publications whose objective addressed therapeutic guidelines to assist in the self-management of patients with active symptoms of major depressive disorder

Exclusion criteria

- Population, concept, and context: publications without information on the population, concept, and context of interest for this study
- Type of study: duplicate studies, opinion articles, letters to the editor, abstracts of conference proceedings, and studies with unavailable text
- Language: studies in languages other than those selected
- Results: publications that did not fit the objectives and did not contain information related to the population, concept, and context of this study

Study Identification and Selection

The process of identifying and selecting studies is one of the most important stages in the conduct of a scoping review. To minimize potential bias, the *JBI Manual for Evidence Synthesis* guidelines were used and explained in a meeting with 6 researchers to ensure that this process was clear to everyone.

Thus, to maintain the rigor of the screening process, the studies obtained from each of the databases based on the eligibility criteria were exported to a reference management software (EndNote; Clarivate Analytics), and duplicates were removed. The remaining studies were then imported into the Rayyan application (Qatar Computing Research Institute) for analysis and selection based on titles and abstracts.

The titles and abstracts were read and analyzed by 2 independent reviewers to identify those that were potentially eligible. After reading the titles and abstracts, the preselected studies were subjected to data analysis and mapping, which consisted of carefully reading and classifying the texts from which the results were extracted.

The selected studies were read in full by 2 additional reviewers (MdPSdSN and PdCT) to confirm their relevance to the research question. Any doubts about the inclusion of studies would be resolved through discussion with a third reviewer until a consensus was reached. It was not necessary to contact a third reviewer.

Data Extraction and Analysis

As recommended by the JBI, the authors developed a tool to extract the data for the scoping review. Those responsible for data extraction (MdPSdSN, PdCT, and PMM) read and carefully classified the texts. The data were documented in a Microsoft Excel (Microsoft Corp) spreadsheet for better data extraction and evaluation. The final studies were determined from this step. The following variables were extracted from the studies: symptoms, authors, year, reference, study title, objective, study design, and mapping of results.

The focuses of the studies were analyzed, and the results and discussion are presented descriptively and quantified as frequencies in tables. The Discussion section synthesizes the evidence found during the review to explore it and compare it with the existing literature.

The mapped results were categorized into 7 major groups and subgroups of therapeutic orientations for the self-management of major depressive disorder based on the similarities between them. For example, therapeutic orientations focused on physical activity or regular exercise and healthy eating were classified as adoption of healthy habits, whereas listening to music, watching movies, reading, painting, going out with friends, taking outdoor walks, and practicing spiritual and religious activities were grouped in the category of leisure or pleasurable activities. This was done for all the therapeutic orientations identified considering the objective of this scoping review.

Regarding the methodological assessment (although it was not mandatory), each included study had its level of evidence identified based on the JBI appraisal tools [13] and the study design.

Thus, classification I was assigned to systematic reviews and meta-analyses of randomized clinical trials, classification II was assigned to randomized clinical trials, classification III was assigned to nonrandomized controlled trials, classification IV was assigned to case-control or cohort studies, classification V was assigned to systematic reviews of qualitative or descriptive

XSL•FO
RenderX

studies, classification VI was assigned to qualitative or descriptive studies, and classification VII was assigned to opinions of authorities or expert committee reports. This hierarchy classifies levels I and II as strong, levels III to V as moderate, and levels VI to VII as weak.

Results

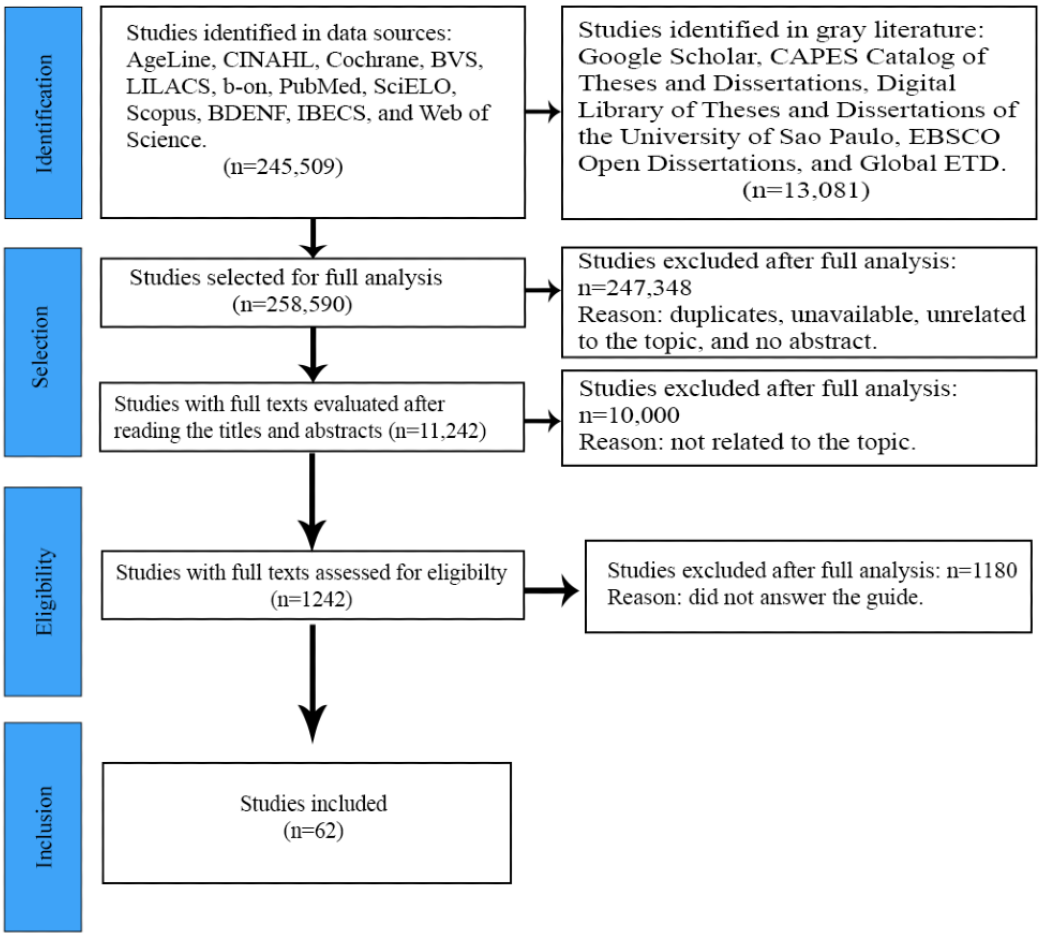
Overview of the Studies Found

Initially, 258,509 potentially eligible publications were found in the databases and 13,081 publications in the gray literature sources, of which 1538 (0.63%) focused on sleep disorders, 8315 (3.39%) focused on appetite disorders, 6732 (2.74%) focused on weight disorders, 5814 (2.37%) focused on feelings of guilt, 2509 (1.02%) focused on irritable mood, 200,002 (81.46%) focused on low self-esteem, 1848 (0.75%) focused on anhedonia, and 31,832 (7.64%) focused on fatigue.

The inclusion and exclusion criteria were applied, and 247,348 duplicate studies were removed, leaving 11,242 (4.55%). Next, a new screening was carried out in search of available publications that had abstracts and dealt with therapeutic guidelines. Of the remaining 11,242 studies, 10,000 (88.95%) were excluded, leaving 1242 (11.05%).

Once the 1242 studies had been selected, the second full reading was carried out. Of the 1242 studies, 1180 (95.01%) were excluded because they did not answer the guiding question. Thus, 62 studies were included in the final sample of this review. Figure 1 shows the results using the PRISMA-ScR flow diagram.

Figure 1. Flow diagram of the search and selection of studies according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 model.



Characteristics of the Included Studies

Of the 62 studies dealing with therapeutic guidelines (Table 1), 71% (n=44) came from indexed data sources, and 29% (n=18) were gray literature indexed on Google Scholar (n=13, 21%), doctoral theses (n=3, 5%), and Master's dissertations (n=2, 3%).

It was found that the studies were published between 2011 and 2023, with a higher number of publications in the years 2018 to 2023 (40/62, 63%), demonstrating that the use of therapeutic guidelines is recently evolving. In terms of language, most publications were in English (North America—United States) and Portuguese (South America—Brazil), totaling 63% (40/62) of the publications (Table 1).

Regarding study design, the studies were clinical trials (30/62, 48%), systematic reviews (6/62, 10%), narrative literature reviews (7/62, 11%), systematic reviews and meta-analyses (6/62, 10%), qualitative studies (3/62, 5%), pilot studies (3/62, 5%), experimental studies (2/62, 3%), case reports (2/62, 3%), noncontrolled studies (1/62, 2%), cross-sectional studies (1/62, 2%), and scoping reviews (1/62, 2%; Table 1).

XSL•FO
RenderX

Table 1. General characterization of the included studies (N=62).

Category and variable	Studies, n (%)
Year of publication	
2011-2017	22 (37)
2018-2023	40 (63)
Origin of the studies	
America	40 (63)
Europe	10 (17)
Asia	9 (15)
Oceania	2 (3)
Africa	1 (2)
Study design	
Clinical trial or experimental, pilot, qualitative, or quantitative study	38 (61)
Systematic review and meta-analysis, scoping review, or literature review	20 (32)
Noncontrolled design, case study, or cross-sectional study	4 (7)

Summary of the Extracted Variables

Table 2 presents the variables extracted from the studies: symptoms, authors, year of publication, reference, title of the study, objective, and study design. All 62 studies focused on people aged >18 years with a diagnosis of major depressive disorder and active symptoms, with a maximum age of 70 years. Of the 15 publications selected on the symptom of sleep disorders, 10 (67%) were published on PubMed [14-23], 1 (7%) was published on SciELO [24], 1 (7%) was published on AgeLine [25], and 3 (20%) were published as gray literature [26-28].

XSL•FO
RenderX

Table 2. Studies found according to symptom, study title, objective and study design.

Symptom and study	Study title	Objective	Study design
Sleep disorders; Yang [26], 2019	"Chinese medicine for major depressive disorder: clinical evidence, patients' experience and expectations, and doctors' perceptions"	To evaluate and understand the evidence on Chinese medicine for major depressive disorder in the literature	Randomized clinical trial; level II
Sleep disorders; Geoffroy et al [14], 2018	"Insomnia and hypersomnia in major depressive episode: prevalence, sociodemographic characteristics and psychiatric comorbidity in a population-based study"	Examine the frequency of sleep complaints, co-occurrences, sociodemographic characteristics, and psychiatric comorbidities associated with each type of sleep profile	Randomized clinical trial; level II
Sleep disorders; Leggett et al [15], 2018	"Bright light as a preventive intervention for depression in late-life: a pilot study on feasibility acceptability and symptom improvement"	Examine the feasibility and acceptability of a portable bright light and its impact on sleep disorders and symptoms of depressive disorders in older adults	Pilot study; level II
Sleep disorders; Luca et al [16], 2013	"Sleep disorders and depression: brief review of the literature, case report, and nonpharmacologic interventions for depression"	Review and discussion of sleep disorders during major depressive disorder (in particular night terrors, nightmares, hypersomnia, and insomnia)	Case report; level VI
Fatigue; Hulme et al [29], 2018	"Fatigue interventions in long term, physical health conditions: a scoping review of systematic reviews"	Map effective guidelines in all clinical conditions for the symptom of fatigue and whether these guidelines can be applied to other symptoms	Scoping review; level I
Sleep disorders; de Sousa Ibiapina et al [24], 2022	"Effects of music therapy on symptoms of anxiety and depression in adults diagnosed with mental disorders: a systematic review"	Identify and synthesize the evidence from studies that have evaluated the effects of music therapy on the symptoms of anxiety and depression in adults with mental disorders	Systematic literature review; level I
Fatigue; Macnamara et al [30], 2018	"Personalized relaxation practice to improve sleep and functioning in patients with chronic fatigue syndrome and depression: study protocol for a randomised controlled trial"	Facilitate new approaches in clinical practice and more meaningful results for patients experiencing states of chronic fatigue	Randomized clinical trial; level II
Anhedonia; Ito et al [31], 2019	"Augmentation of positive valence system-focused cognitive behavioral therapy by inaudible high-frequency sounds for anhedonia: a trial protocol for a pilot study"	Testing the effect of increasing inaudible high-frequency sounds on the effectiveness of CBT[a] focused on the positive valence system to treat anhedonia	Pilot study; level II
Anhedonia; Braun et al [32], 2019	"A pilot study investigating the effect of music-based intervention on depression and anhedonia"	To investigate the effect of a music therapy-based approach to managing major depressive disorder and its associated symptoms.	Clinical trial; level III
Anhedonia; Ebrahem and Masry [33], 2017	"Effect of relaxation therapy on depression, anxiety, stress and quality of life among diabetic patients"	To evaluate the effect of relaxation therapy on major depressive disorder and its associated symptoms such as anxiety, stress, quality of life and blood glucose levels.	Quasi-experimental study; level II
Low self-esteem; Kinser et al [34], 2013	"'A feeling of connectedness': perspectives on a gentle yoga intervention for women with major depression"	To evaluate the feasibility, acceptability, and effects of yoga among women with major depressive disorder	Randomized clinical trial; level II
Low self-esteem; Grandia [35], 2014	"Patient-initiated strategies for self-management of depression and low mood: understanding theory and changing behavior"	Examine the theory of patient-planned behavior for major depressive disorder self-management	Randomized clinical trial; level II
Sleep disorders; Prathikanti et al [25], 2017	"Treating major depression with yoga: a prospective, randomized, controlled pilot trial"	Examine a therapeutic guideline of hatha yoga for 8 weeks as monotherapy for major depressive disorder	Randomized clinical trial; level II
Guilt; Sathyanarayan et al [36], 2019	"Role of yoga and mindfulness in severe mental illnesses: a narrative review"	To review the effectiveness of yoga and mindfulness as a treatment modality for major depressive disorder	Literature review; level V
Low self-esteem; van Grieken et al [37], 2015	"Patients' perspective on self-management in the recovery from depression"	Determine therapeutic guidelines for patients to recover from major depressive disorder	Randomized clinical trial; level II
Sleep disorders; Ma et al [17], 2018	"The effects of Tai Chi on sleep quality in Chinese American patients with major depressive disorder: a pilot study"	Evaluate the effects of tai chi on sleep quality and functioning among patients with major depressive disorder	Pilot study; level II

XSL•FO
RenderX

Symptom and study	Study title	Objective	Study design
Guilt; Sarsak et al [38], 2020	"Applied occupational therapy for major depressive disorder: clinical case report"	Describe an intervention carried out via occupational therapy on an older woman with major depressive disorder and suicidal thoughts	Clinical case study; level VI
Sleep disorders; Xu et al [18], 2023	"Clinical evidence for association of acupuncture with improved major depressive disorder: a systematic review and meta-analysis of randomized control trials"	To determine the efficacy and safety of acupuncture for major depressive disorder	Meta-analysis; level I
Angry mood; Souza [39], 2019	"Chinese auricular acupuncture in the treatment of depression"	To investigate the efficacy of Chinese auricular acupuncture for major depressive disorder	Study of a qualitative nature; level V
Sleep disorders; Farrar and Farrar [19], 2020	"Clinical aromatherapy"	Investigate the history, supporting theories, guidelines, plant sources, safety, and pathophysiological responses of clinical aromatherapy in nursing	Systematic literature review; level I
Guilt; Birgitta et al [40], 2018	"Treatment of depression and/or anxiety - outcomes of a randomised controlled trial of the tree theme method® versus regular occupational therapy"	To compare therapeutic guidelines for major depressive disorder in relation to activities of daily living, psychological symptoms of major depressive disorder, and health-related aspects	Randomized clinical trial; level II
Sleep disorders; Ter Heege et al [20], 2020	"The clinical relevance of early identification and treatment of sleep disorders in mental health care: protocol of a randomized control trial"	Identify the prevalence of sleep disorders in different mental disorders, including major depressive disorder	Meta-analysis; level I
Sleep disorders; Russi et al [27], 2022	"Physical therapy intervention in the treatment of insomnia"	To evaluate a tool for restoring physiological balance and improving sleep in major depressive disorder	Experimental study; level III
Guilt; Duggal [41], 2019	"Self-management of depression: beyond the medical model"	Describe a new self-management paradigm in major depressive disorder centered on the patient beyond the improvement of clinical symptoms	Systematic literature review; level I
Weight change; Naslund et al [42], 2017	"Lifestyle interventions for weight loss among overweight and obese adults with serious mental illness: a systematic review and meta-analysis"	Evaluate the effects of a therapeutic approach to lifestyle changes on short- and long-term changes in body weight in people with major depressive disorder	Systematic review and meta-analysis; level I
Angry mood; de Cassia Rondina et al [43], 2018	"Practicing physical exercise and symptoms of depression in college students"	To investigate the relationship between major depressive disorder symptoms and physical activity patterns in university students	Randomized clinical trial; level II
Weight change; Daumit et al [44], 2013	"A behavioral weight-loss intervention in persons with serious mental illness"	Determine the effectiveness of guidelines for weight loss in adults with severe mental illness, including major depressive disorder	Essay randomized clinical trial; level II
Appetite change; Kazemi et al [45], 2020	"Effect of probiotic and prebiotic vs placebo on psychological outcomes in patients with major depressive disorder: a randomized clinical trial"	Investigate the effect of supplementation with probiotics and prebiotics on appetite, BMI, weight, and energy intake in patients with major depressive disorder	Randomized clinical trial; level II
Weight change; Colombari et al [46], 2018	"The effect to behavioral weight-loss intervention on depressive symptoms among Latino immigrants in a randomized controlled trial"	Testing the effectiveness of therapeutic guidelines for changing lifestyle life versus usual care among adults with obesity and major depressive disorder	Systematic review and meta-analysis; level I
Weight change; Carraça et al [47], 2013	"The association between physical activity and eating self-regulation in overweight and obese women"	To assess the importance of physical exercise and healthy eating for managing low self-esteem and depressed mood in people with major depressive disorder	Randomized clinical trial; level II
Sleep disorders; Rethorst et al [21], 2015	"IL-1β and BDNF are associated with improvement in hypersomnia but not insomnia following exercise in major depressive disorder"	Examine changes in hypersomnia and insomnia after increased exercise in people with major depressive disorder	Randomized clinical trial; level II
Angry mood; de Vaz Pato Oom [48], 2019	"Physical activity in the treatment of young adult depressive disease"	To evaluate the beneficial effect of physical activity in the treatment of major depressive disorder	Literature review; level V

XSL•FO
RenderX

Symptom and study	Study title	Objective	Study design
Low self-esteem; Kandola et al [49], 2019	"Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity"	Evaluate the main biological and psychosocial mechanisms of activity as an antidepressant	Systematic review; level I
Low self-esteem; Barton et al [50], 2012	"Exercise-, nature- and socially interactive-based initiatives improve mood and self-esteem in the clinical population"	Evaluate initiatives to promote health through outdoor exercise involving a population with major depressive disorder	Randomized clinical trial; level II
Weight change; Soares et al [51], 2020	"Effects of physical exercise on obesity and depression: a review"	To integrate the findings on the effects of physical exercise on obesity and major depressive disorder	Literature review; level V
Anhedonia; Brush et al [52], 2022	"A randomized trial of aerobic exercise for major depression: examining neural indicators of reward and cognitive control as predictors and treatment targets"	To demonstrate the effectiveness of aerobic exercise among adults with major depressive disorder	Randomized controlled clinical trial; level II
Anhedonia; Belvederi et al [53], 2019	"Physical exercise in major depression: reducing the mortality gap while improving clinical outcomes"	Provide a concise update on the effectiveness of physical exercise on major depressive disorder and reduction in cardiovascular mortality	Systematic literature review; level I
Anhedonia; Turner et al [54], 2019	"Physical activity and depression in MS: the mediating role of behavioral activation"	Evaluate the impact of physical activity to improve fatigue and symptoms of depressive disorders in individuals with multiple sclerosis	Randomized controlled clinical trial; level II
Anhedonia; Toups et al [55], 2017	"Exercise is an effective treatment for positive valence symptoms in major depression"	To evaluate the effect of physical exercise on the symptoms of major depressive disorder	Essay randomized clinical trial; level II
Anhedonia; Archer et al [56], 2014	"Effects of physical exercise on depressive symptoms and biomarkers in depression"	To investigate the effects of physical exercise on depressive symptoms and biomarkers in major depressive disorder	Systematic review of the literature; level I
Angry mood; Bains and Abdijadid [57], 2022	"Major depressive disorder"	Identify, describe, and review the etiology, management, and clinical presentation of major depressive disorder	Review and meta-analysis; level I
Low self-esteem; Bajaj et al [58], 2016	"Mediating role of self-esteem on the relationship between mindfulness, anxiety, and depression"	To examine the mediating effects of self-esteem on the association among mindfulness, anxiety, and major depressive disorder	Cross-sectional study; level VI
Low self-esteem; Randal et al [59], 2015	"Mindfulness and self-esteem: a systematic review"	Evaluate studies investigating the association between mindfulness and self-esteem in relation to major depressive disorder	Systematic literature review; level I
Angry mood; Farb et al [60], 2018	"Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy"	Evaluate relapse rates in patients with depression in remission receiving CBT based on mindfulness and CT[b]	Meta-analysis; level I
Sleep disorders; Kuyken et al [22], 2019	"Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials"	Conduct a meta-analysis of individual patient data to examine the effectiveness of mindfulness-based CT in the treatment of major depressive disorder	Systematic review and meta-analysis; level I
Appetite change; Katterman et al [61], 2014	"Mindfulness meditation as an intervention for binge eating, emotional eating, and weight loss: a systematic review"	Examine studies on mindfulness for the treatment of binge eating associated with major depressive disorder	Systematic review of the literature; level I
Guilt; Fawns [62], 2013	"Mindfulness based cognitive therapy and emotion focused therapy as treatments for major depression"	To describe the symptomatology and treatment options for major depressive disorder, and to examine and compare the effectiveness of two treatment orientations: mindfulness-based cognitive therapy and emotion-focused therapy	Literature review; level V
Angry mood; Economides et al [63], 2018	"Improvements in stress, affect and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial"	To evaluate the mindfulness-based Headspace smartphone app for managing symptoms of major depressive disorder	Randomized clinical trial study; level II
Appetite change; Person [64], 2021	"A two-study investigation into the link between rumination and night eating, and symptom improvement following a mindfulness-based intervention"	To explore and investigate the link between night eating syndrome and depressive symptoms	Clinical study; level II

XSL•FO
RenderX

Symptom and study	Study title	Objective	Study design
Sleep disorders; Bellingham [28], 2019	"Spiritually focused mindfulness meditation: an interpretative phenomenological analysis of the effect of spiritually focused mindfulness meditation on depression with a clinical population"	Explore the use of focused mindfulness meditation for major depressive disorder	Qualitative study; level VI
Guilt; Schanche et al [65], 2020	"The effects of mindfulness-based cognitive therapy on risk and protective factors of depressive relapse - a randomized wait-list controlled trial"	Explore the effects of mindfulness-based CT on risk and protective factors for major depressive disorder relapse in the domains of cognition, emotion, and self-relationship	Randomized clinical trial study; level II
Angry mood; Rech et al [66], 2022	"Techniques for managing the emotion of anger: a systematic review"	Identify the main techniques for managing anger in adults with major depressive disorder	Systematic review of the literature; level I
Sleep disorders; Chung et al [23], 2020	"Mobile app use for insomnia self-management in urban community-dwelling older Korean adults: retrospective intervention study"	Explore the relationship between sleep quality, memory concerns (memory loss) and depressive symptoms	Randomized clinical trial; level II
Low self-esteem; Moloud et al [67], 2022	"Cognitive-behavioral group therapy in major depressive disorder with focus on self-esteem and optimism: an interventional study"	Determine the effect of group CBT for the management of low self-esteem in patients with major depressive disorder	Randomized clinical trial study; level II
Guilt; Dobkin et al [68], 2019	"Cognitive behavioral therapy improves diverse profiles of depressive symptoms in Parkinson's disease"	Examine the impact of CBT on different depressive symptoms in Parkinson disease	Randomized clinical trial; level II
Angry mood; Santos [69], 2017	"Efficacy of procedural cognitive therapy and behavioral activation in the treatment of major depressive disorder: a randomized clinical trial"	To compare the effectiveness of psychotherapies for patients with major depressive disorder	Randomized clinical trial study; level II
Angry mood; Ahern et al [70], 2017	"Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis"	Evaluate the clinical effectiveness and evidence for the use of online CBT as an affordable treatment solution for major depressive disorder	Systematic review and meta-analysis; level I
Low self-esteem; Korrelboom et al [71], 2012	"Competitive memory training (COMET) for treating low self-esteem in patients with depressive disorders: a randomized clinical trial"	Evaluate whether competitive memory training is an effective intervention for patients with major depressive disorder	Randomized clinical trial study; level II
Anhedonia; Wang et al [72], 2020	"Guided self-help behavioral activation intervention for geriatric depression: protocol for pilot randomized controlled trial"	Pilot a therapeutic guided self-help intervention for the treatment of major depressive disorder in older adults	Randomized clinical trial; level II
Guilt; Alves and Bonvicini [73], 2022	"The role of behavioral activation in the management of depressive symptoms"	Describe and evaluate behavioral activation as a psychotherapy tool and guideline for managing mood, behavior, and emotions in patients with major depressive disorder	Systematic review; level I
Fatigue; Wisenthal et al [74], 2019	"Insights into cognitive work hardening for return-to-work following depression: qualitative findings from an intervention study"	Contribute to the literature on the effectiveness of cognitive maturation to return to work after episodes of major depressive disorder	Randomized clinical trial; level II
Weight changes; Berman et al [75], 2016	"Uncontrolled pilot study of an acceptance and commitment therapy and health at every size intervention for obese, depressed women: accept yourself!"	Evaluate the feasibility and outcomes of a new treatment based on self-acceptance for women with obesity and major depressive disorder	Randomized uncontrolled clinical trial; level II

[a]CBT: cognitive behavioral therapy.

[b]CT: cognitive therapy.

For the symptom of anhedonia, 9 publications were selected: 3 (33%) published on BVS [32,52,72], 3 (33%) published on PubMed [53-55], 2 (22%) published on Scopus [31,56], and 1 (11%) published as gray literature [33]. Similarly, the symptom of low self-esteem was identified in 9 publications: 5 (56%) published in PubMed [34,37,49,50,71], 1 (11%) published on Scopus [58], and 3 (33%) published as gray literature [35,59,67]. Furthermore, 9 publications were also selected for the symptom of irritated mood: 4 (44%) published on PubMed [57,60,63,70] and 5 (56%) published as gray literature [39,43,48,66,69].

For the symptom of guilt, 8 publications were selected: 4 (50%) published on PubMed [40,41,65,68] and 4 (50%) published as gray literature [36,38,62,73]. Regarding the symptom of weight change, 6 publications were identified: 5 (83%) published on PubMed [42,44,46,47,75] and 1 (17%) published as gray

XSL•FO
RenderX

literature [51]. For the symptom of fatigue, 3 publications were selected: 2 (67%) published on PubMed [29,30] and 1 (33%) published on CINAHL [74]. Finally, for the symptom of appetite change, 3 publications were identified: 2 (67%) published on PubMed [45,61] and 1 (33%) published as gray literature [64]. Regarding the level of evidence, 48% (30/62) of the studies were level II, 34% (21/62) were level I, 8% (5/62) were level V, 6% (4/62) were level VI, and 3% (2/62) were level III.

Self-Management of Major Depressive Disorder

The evidence available from the included studies to promote self-management of the symptoms of major depressive disorder describes the characteristics, benefits, and applicability of therapeutic guidelines in everyday life, demonstrating that these guidelines can promote a safe space to share conversations on sensitive subjects such as stigma and prejudice, adverse drug events, diagnosis, and relapse [14-75].

The findings also point out that, to promote self-management through therapeutic guidelines, it is necessary to know the functionality of the patient or family, have access to community resources and a social support network, and monitor the participation of these individuals in their care [16,33,35,37,40]. The interdisciplinary team's online or face-to-face follow-up of patients with major depressive disorder involves monitoring the use of therapeutic guidelines and whether they are being carried out correctly and appropriately considering the individual's uniqueness, as well as monitoring changes in symptoms and problems with medication, providing social support, and identifying patients at high risk of relapse [23,41].

Another common point cited in the studies was the ability of health professionals to instruct patients with major depressive disorder on how to deal with symptoms and biosociopsychological demands [16,31-33]. Professionals require skills and good communication with patients when instructing them on the use of therapeutic guidelines for managing the disorder, since incorrect instructions or misunderstandings may make the application of therapeutic guidelines less efficient [23,35,37,40,67,70,74,75]. If the professional advises the patient to sunbathe (phototherapy) during the day, but does not specify the time, this may cause the patient to take this guidance at inappropriate times and be harmful to their health. Therefore, phototherapy would not be effective. The findings encourage a multidisciplinary approach to the applicability of therapeutic guidelines, focusing on the specificities of each person and collective actions, strengthening the patient's ability to self-manage.

Therapeutic Guidelines for Self-Management

The analysis of the studies, as described previously, resulted in the mapping and categorization of 7 large groups of therapeutic guidelines for the self-management of major depressive disorder. These guidelines are made up of a set of actions or activities (n=40), which have been called subinterventions based on the similarities between them (Table 3).

In total, 52% (32/62) of the therapeutic guidelines identified fell into the category of psychotherapy, which helps monitor and plan activities with the person (in the context of their symptoms), manage adverse experiences, and develop social skills [14,16,17,22,23,26,28,29,31,36-38,46,56,58-72,74,75]. Adoption of healthy habits (25/62, 40%) and integrative and complementary practices (17/62, 27%) were the second and third most observed categories. The process of adopting integrative and complementary practices [17,19,20,24-26,29-40], together with healthy habits such as eating a balanced diet and exercising frequently and for the right duration, showed a good response in improving depressive conditions, as well as being an important factor in preventing relapses of major depressive disorder [21,26,27,29,30,33,37,38,41-57].

Finally, relaxation techniques (9/62, 14%), consultation with a health professional (14/62, 22%), pharmacological therapy (9/62, 14%), and leisure or pleasurable activities (4/62, 6%) were the least observed therapeutic guidelines in the studies. However, they offered promising insights into self-management of the disorder and underscored the importance of a multifaceted approach to managing this complex condition [16,20,23,26-30,33,36-38,41,50,57,66].

It is widely recognized that the concept of self-management is traditionally associated with approaches that the patient can carry out autonomously, such as adopting healthy habits, physical exercise, and relaxation techniques. However, it is crucial to emphasize that, in the context of major depressive disorder, self-management transcends these independent symptom management practices [37].

It involves empowering the patient to take active control and gain an in-depth understanding of their condition through the application of a variety of therapeutic interventions [35]. Patients with major depressive disorder may experience feelings of frustration and a range of negative thoughts, see everyday problems as major catastrophes, have difficulty recognizing efforts that awaken hope for life, and cope with the possible adverse events that psychiatric medications can cause [66].

For this reason, orientation toward psychotherapy and psychopharmaceuticals is justified as a self-management strategy as a way of coping with the adverse effects of medication and all the feelings of self-demand, excessive demands, and frustration caused by the disorder.

In relation to the population to which this study refers, the results show that patients with major depressive disorder have motivational, cognitive, and psychological deficits that can modify their ability to cope and reason [22,60,63,65].

In view of this, self-management of the signs and symptoms resulting from major depressive disorder needs to take place at the beginning of treatment, with the support of family or caregivers; friends; community organizations; and, most especially, the multi-professional health team to help with the correct use of therapeutic options until the person is able to self-manage independently [14-75]. Therefore, the therapeutic guidelines found in this review are not restricted to the professional-patient relationship; these relationships are broad and include families and the community.

XSL•FO
RenderX

Table 3. Mapping of the results regarding the therapeutic guidelines.

Therapeutic guideline	Subinterventions
Adopting healthy habits	• Physical activity or regular exercise [21,26,27,29,30,33,37,38,41-57] • Healthy eating [26,27,29,37,41-46]
Leisure or pleasurable activities	• Listening to music, watching movies, reading, painting, going out with friends, taking outdoor walks, and practicing spiritual and religious activities [37,38,41,50]
Relaxation techniques	• Progressive muscle relaxation [33,41,66] • Respiratory rehabilitation [16,26,27,29,30,36] • Stretching [16,26,27,29,41,66] • Warm baths [16] • Massages [29,41,66] • Meditation [30]
Pharmacological therapy	• Psychopharmacological therapy [16,20,26,28,29,36,38,41,57]
Consultation with a health professional	• Psychoeducation or professional counseling [16,20,23,26,27,29,36-38,41-43,57,66]
Psychotherapy	• Behavioral activation [69,73] • Cognitive behavioral psychotherapy [14,16,23,26,29,31,38,58,67-71] • Mindfulness [22,26,28,36-38,59-66] • Psychotherapy focused on problem-solving [29,38,74] • Competitive memory training psychotherapy [71] • Self-efficacy psychotherapy [46] • Guided self-help psychotherapy [23,72] • Interpersonal psychotherapy [17,23,56] • Therapy focused on acceptance, commitment [62], and emotion [75]
Integrative and complementary practices	• Acupuncture [19,26,29,39] • Music therapy [24,30-33] • Phytotherapy [26,29] • Phototherapy [26,29] • Aromatherapy [19] • Cryotherapy [29] • Thermotherapy [29] • Tai chi [17,29,38] • Reflexology [29] • Qigong [29] • Yoga [25,29,34-37] • Art therapy [40] • Chromotherapy [20]

Discussion

Principal Findings

From this review, it was possible to identify the therapeutic guidelines for the self-management of the symptoms of major depressive disorder in scientific production as psychotherapy, adoption of healthy habits, integrative and complementary practices, relaxation techniques, consultation with a health professional, pharmacological therapy, and leisure or pleasurable activities.

The selected studies show that knowledge on the subject significantly improves self-management of the symptoms of major depressive disorder through these therapeutic guidelines to promote behavior changes and awareness of the symptoms of the disorder, prevention of relapses, reduction of the perception of obstacles, and increased adherence to treatment.

A small number of articles published in journals on the subject were found for all the symptoms investigated, especially irritable mood, followed by fatigue, weight change, appetite change, sleep changes, and low self-esteem, necessitating the addition of secondary documents that considerably broadened the identification of other therapeutic guidelines.

Most of the studies originated in America (40/62, 63%), where the largest number of people with major depressive disorder is concentrated. Regarding the concept that triggered the research, the studies involved health professionals, students, and the general population, covering the definition, description, comparative analysis, efficiency, and applicability of therapeutic guidelines. These studies were included to increase the possibility of nurses playing a leading role in the self-management of symptoms of major depressive disorder [14-75].

The factors considered important for the self-application of therapeutic guidelines in the studies were cognition; level of education; environmental factors; and the functional capacity of the patient, caregivers, and family members. It should be emphasized that, for self-management, clients who are going

XSL•FO
RenderX

to follow these therapeutic guidelines should receive them in clear and objective language and format, with specificity and relevance for each symptom, enabling adherence to treatment and favoring the therapist–patient or caregiver bond [26]. These therapeutic guidelines are also considered low cost and easy to use, optimize rehabilitation and autonomy, and encourage self-care [26]. It is important to highlight the definitions of each of the therapeutic guidelines identified in this review.

Integrative and complementary practices, as therapeutic resources based on traditional knowledge, were the therapeutic guidelines most commonly found in the literature as effective for the self-management and management of major depressive disorder symptoms, and their applicability aimed to induce a state of harmony and balance throughout the body.

Phototherapy has been shown to reduce depressive symptoms and improve irregularities in sleep patterns and quality, especially in older adults with major depressive disorder, as long as it is used with a certain frequency and is administered before 10 AM [14-16,26]. It has also been used in the control of nocturnal hyperphagia and has shown reductions in the percentage of food eaten after dinner and in the number of nocturnal meals eaten per week, reducing the rate of overweight and obesity, as well as being used in the control of fatigue, where it has shown a reduction in the feeling of tiredness [29].

Music therapy applied to adults [24] has been shown to be effective in physical and mental relaxation, reducing anxious and depressive symptoms and promoting well-being in a conscious and healthy way [30]. Listening to soothing noises and classical music associated with rhythmic breathing produced a significantly higher rate of adherence to major depressive disorder treatment, especially in the patients with the highest severity [31-33].

A study has shown that anhedonia and clinically significant depressive symptoms can be resistant to standard treatment but adjuvant treatment with high-frequency inaudible sound therapy increases the reward of related brain circuits and has a synergistic effect on anhedonia [31]. Integrating music therapy with conventional major depressive disorder treatment (therapies and medication) gives people the opportunity to get in touch with their emotions and provides distraction and a means of communication capable of overcoming barriers and limits to verbal expression [30].

Yoga [29,34,35] has been a beneficial exercise for managing sleep quality and other symptoms of major depressive disorder [25,36]. It is an important tool for promoting self-care and, consequently, care for everything around the patient [37]. A study carried out to explore the experiences of patients with major depressive disorder showed that practicing yoga as a self-management strategy helped them gain better insights into their own condition, improving the quality of their health care [37].

Tai chi has been shown to be more effective than traditional rehabilitation (psychotropic drugs and therapies) in relation to insomnia, with an improvement in the quality, duration, and efficiency of sleep and a reduction in thoughts of guilt associated with major depressive disorder [17,29,38]. Acupuncture has been shown to be effective in treating people with major depressive disorder [18,39], mainly in reducing sleep disorders but also in symptoms related to changes in appetite, irritability, anxiety, and sadness, and has not been shown to be effective for suicidal ideation [26,29].

With regard to herbal medicines, St John's wort, *chai hu*, and *gancao* were the most widely used. In 3% (2/62) of the studies, after 2 weeks of administering these herbs (alone or in combination) as an alternative to conventional antidepressants, they produced a statistically significant improvement in symptoms of altered sleep [26,29]. Although the effects of these herbs are not fully understood, it is likely that they produce antidepressant effects through multiple pathways or targets that interact with each other.

To a lesser extent among the studies but no less importantly, aromatherapy [19], cryotherapy, thermotherapy, reflexology, qigong [29], art therapy [40], and chromotherapy [20] improved empowerment in the search for self-care and taking responsibility for one's own health, as well as reducing levels of anxiety and stress, improving sleep disorders, improving the immune system, and lowering blood pressure levels [29].

People with major depressive disorder have shown improvements in sleep disorders, anxiety, and stress, as well as enhancements in immune system function and reductions in blood pressure levels when using respiratory rehabilitation [16,26,27,29,30,36], stretching [16,26,27,29,41,66], warm baths [16], massages [29,41,66], acupuncture, herbal medicine, phototherapy, reflexology, qigong, and yoga [17,19,20,24,26,29,34,38-40] as therapeutic guideline.

Cryotherapy and thermotherapy are not commonly used to treat the symptoms of major depressive disorder, but they help relax the body and mind as a physiotherapeutic resource [16,29]. There was a significant improvement in stress and body pain related to fatigue, anxiety, and swelling in the feet and legs [29].

Another important therapeutic guideline was the adoption of healthy lifestyle habits, such as healthy eating [17,26,27,29,41-46] and physical exercise [21,26,27,30,33,37,38,41-57]. These reduce stress; improve mood, body image perception, and self-esteem; stimulate cognitive functioning; promote greater satisfaction with life [45,46]; and can help people feel stronger and more capable, helping reduce the feelings of hopelessness and helplessness associated with major depressive disorder [53,54].

It should be noted that, regardless of the physical activity chosen, people with major depressive disorder need to make it part of their routine and do it [47] according to their tolerance and state of health [54]. An important concern in relation to physical activity for people with major depressive disorder is the fact that some of the common symptoms of depression (fatigue, lack of energy, psychomotor retardation, despair, and feelings of worthlessness) interfere with the motivation to exercise [57].

In more severe cases, practicing physical activity can be difficult for people with major depressive disorder [55]. This can compromise adherence and long-term permanence in the exercise program [52]. It is recommended to prescribe structured

XSL•FO
RenderX

exercises based on the activities already practiced by the patient to reduce obstacles [33].

Another group of therapeutic guidelines that were also commonly identified in this review were psychotherapies. Mindfulness-based therapy [58,59] is a promising therapeutic approach for treating major depressive disorder and is part of the list of integrative and complementary practices [26,37,60]. It combines elements of therapeutic guidelines [22,61] with the aim of helping people develop a full awareness of the present moment, accepting and acknowledging their thoughts, emotions, and bodily sensations without judgment [62] and learning to observe their negative thoughts and dysfunctional thought patterns without allowing themselves to be involved with or controlled by them [63].

Mindfulness improves self-management of prodromal symptoms [64], represents a stability factor when practiced frequently and in a supervised manner [28], positively influences memory and distorted feelings of guilt, increases cognitive resilience, balances mood and sleep, reduces high levels of stress, and improves self-esteem and anxiety [65] associated with major depressive disorder relapse.

Similarly, cognitive behavioral psychotherapy [26,66] uses positive reinforcement techniques and systematic desensitization, helping people cope with stress, face the challenges of everyday life, change negative thought patterns and behaviors to more realistic and positive ones [23,67], and identify the first signs of relapse and prevent it as it focuses on social skill training [68,69]. Recently, online cognitive behavioral psychotherapy has been shown to be effective for depressed and angry mood as a cost-effective treatment modality for major depressive disorder [70].

Other psychotherapies aimed at influencing the person with major depressive disorder and helping them modify emotional, cognitive, and behavioral problems were self-efficacy psychotherapy [46]; competitive memory training [71]; guided self-help psychotherapy [23,72]; behavioral activation [69,73]; interpersonal psychotherapy [38,57,67]; psychotherapy focused on problem-solving [29,38,74]; and therapy focused on acceptance, commitment, and emotion [62,75].

Leisure time or pleasurable activities as a social practice, contrary to what is often thought, are not only carried out during the summer or on vacation but also in between daily obligations [37]. Thus, they can take the form of individual or group dynamics, whether it is reading a book, listening to music, or going for a walk with friends [37,50]. The relationship between leisure and health has fostered new techniques for the treatment of depressive disorders with the aim of enabling the individual's psychic and social readaptation [37,50].

Painting for leisure, for example, can promote the manifestation of feelings through emotions expressed verbally or not, thus helping understand the mind and its sorrows [38]. In this sense, of the possibilities for leisure activities recommended by the professionals who collaborated with the data found, art therapy [40], group and individual outings [41,50], trips [38], reading books and watching movies [37,38], listening to music [31], and attending religious institutions [50] were the most cited in the studies.

In the context of pharmacological treatment, antidepressants associated with psychotherapies [26,29,37] should be administered with caution and under medical supervision [20,38,41] based on the therapeutic alliance, clinical history, monitoring and reassessment of psychiatric conditions, and adequacy of the diagnosis, and guidance should be provided for families [28,57].

In 5% (3/62) of the studies, the association of psychotropic drugs with therapeutic guidelines was highlighted (due to the severity of the major depressive disorder clinical picture) even in the face of difficulties with medication adherence for different reasons [16,29,38], which should be anticipated and addressed proactively in every contact with the patient. Approximately 30% to 40% of people using antidepressants still do not respond as expected to treatment [27,41].

Relaxation techniques appeared in this review as a more appropriate way of coping with the psychological, environmental, and social stimuli that people with major depressive disorder may face. Among these strategies, breathing techniques [16,26,27,29,30,36], progressive muscle relaxation [33,41,66], massages [29,41,66], stretching [16,26,27,29,66], a warm bath [16], and meditation [30] were the most cited.

In the context of major depressive disorder, these techniques promote muscle relaxation, relieve stress and tension, rebalance emotions, produce hormones such as endorphins, provide a sense of well-being, and significantly improve sleep quality [16]. Incorporating these practices throughout the day or even before going to bed helps people let go of tiredness, become aware of the present moment, and pay attention to their own bodies [16,66].

Finally, therapeutic guidelines for psychoeducation or professional counseling can be used to help with self-management as major depressive disorder generates changes in the family system and structure [16,18,26,36,38], resulting in the need for clarification to relieve anxieties and doubts to improve the psychological well-being of the patient [20,37,42,43].

Thus, the wide variety of therapeutic guidelines found can increase knowledge about the symptoms and aspects of major depressive disorder, allowing the person to self-manage symptoms such as fatigue, sleep disturbances, anhedonia, and mood swings [16,27,29,66] in a more responsible and autonomous way.

These therapeutic guidelines are characterized by providing information on the diagnosis, etiology, prognosis, and course of the illness; the identification of early signs of crisis; the importance of adherence to medication or psychosocial treatment; the promotion of healthy habits and regularity in lifestyle (sleep, diet, physical activity, and substance use); and how to deal with the stigmatization, doubts, fears, and myths regarding major depressive disorder [23,41,42].

Therefore, the implications for practice are based on the fact that therapeutic guidelines help in the self-management of major

XSL•FO
RenderX

depressive disorder through behavioral, social, and emotional changes that allow for better construction of clinical reasoning, adaptation, autonomy, coping, and improvement of general health.

Limitations

The limitations of this scoping review are related to the nature of the review itself as its aim was to provide an overview of therapeutic guidelines, which may not be sufficient for the self-management of major depressive disorder as it is a complex clinical condition with a heterogeneity of symptoms. Furthermore, despite efforts to develop a comprehensive search strategy, it was difficult to find controlled and uncontrolled descriptors for the term "therapeutic guidelines," which is not an indexed descriptor.

Conclusions

By analyzing the methodological approach adopted in this study, and in accordance with the proposed objectives, it was possible to highlight a set of therapeutic guidelines to support a person's self-management in the context of major depressive disorder symptoms. This does not preclude further study of their effectiveness in the self-management of major depressive disorder symptoms but demonstrates their contribution to promoting self-management of major depressive disorder and their power to prevent possible complications.

Self-management is a promising strategy that emphasizes the person's responsibility in the care process. It goes beyond participation in health care interactions and includes dealing with symptoms and disability; managing medication and monitoring indicators; maintaining adequate levels of nutrition and exercise; and adjusting to psychological, social, and lifestyle demands.

These therapeutic guidelines can be applied by all members of a multi-professional mental health team, including nurses, provided they are trained as resources to help the person acquire new knowledge, behaviors, and skills for self-management. These guidelines are also an important resource for health professionals in their role as facilitators of a person's development of knowledge and skills to ensure their capacity for self-management.

Acknowledgments

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel, Brazil (finance code 001) and by National Council for Scientific and Technological Development Ministry of Science, Technology and Innovation call 10/2023, Track A—Emerging Groups (process 403740/2023-1). The funders had no role in the study design; data collection, analysis, and interpretation; writing of the manuscript; or decision to publish. Generative artificial intelligence was not used in the writing of this manuscript.

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Authors' Contributions

PdCT contributed to the conceptualization, data curation, investigation, methodology, resource management, validation, and software implementation. PdCT was also responsible for visualization and contributed to both drafting the original manuscript and reviewing and editing it; PMM was responsible for data curation, formal analysis, and investigation; DMP contributed to investigation, methodology, and resource management; MAMJ was responsible for validation and visualization; MAdL contributed to data curation and formal analysis. CSF was involved in data curation, formal analysis, supervision, and drafting the original manuscript; CSVdBS contributed to data curation, formal analysis, and drafting the original manuscript. MdPSdSN contributed to the conceptualization of the study, data curation, funding acquisition, investigation, methodology, project administration, resource management, supervision, validation, and visualization. MdPSdSN also played a key role in drafting the original manuscript and in reviewing and editing it. All authors have read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 85 KB - ijmr_v14i1e63959_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 40 KB - ijmr_v14i1e63959_app2.docx]

References

XSL•FO
RenderX

1. Mônego BG, Fonseca RP, Teixeira AL, Barbosa IG, Souza LC, Bandeira DR. Major depressive disorder: a comparative study on social-emotional cognition and executive functions. Psic Teor e Pesq 2022;38(3):27. [doi: 10.1590/0102.3772e38217.en]
2. Sousa NF, Barros MB, Medina LP, Malta DC, Szwarcwald CL. Association of major depressive disorder with chronic diseases and multimorbidity in Brazilian adults, stratified by gender: 2019 National Health Survey. Rev Bras Epidemiol 2021;24(suppl 2):e210015 [FREE Full text] [doi: 10.1590/1980-549720210015.supl.2] [Medline: 34910069]
3. GHDx: global health data exchange results tool. Institute for Health Metrics and Evaluation. 2019. URL: https://ghdx.healthdata.org/ [accessed 2023-04-30]
4. Bernhardt M, Klauke S, Schröder A. Longitudinal course of cognitive function across treatment in patients with MDD: a meta-analysis. J Affect Disord 2019 Apr 15;249:52-62. [doi: 10.1016/j.jad.2019.02.021] [Medline: 30753954]
5. Saraceno B. Citizenship as a form of tolerance. Rev Ter Ocup Univ São Paulo 2011 Aug 01;22(2):3-101. [doi: 10.11606/issn.2238-6149.v22i2p93-101]
6. Dutra VF, Bossato HR, Oliveira RM. Mediating autonomy: an essential care practice in mental health. Esc Anna Nery 2017;21(3):33. [doi: 10.1590/2177-9465-ean-2016-0284]
7. Bauer M, Whybrow PC, Angst J, Versiani M, Möller HJ. World federation of societies of biological psychiatry (WFSBP) guidelines for biological treatment of unipolar depressive disorders, part 2: maintenance treatment of major depressive disorder and treatment of chronic depressive disorders and subthreshold depressions. World J Biol Psychiatry 2002 Apr 12;3(2):69-86. [doi: 10.3109/15622970209150605] [Medline: 12479080]
8. Nóbrega MD, Tibúrcio PC, Fernandes MC, Fernandes CS, Santos CS, Magalhães BM. Exploring the use of mobile applications for self-management of mental health treatment: scoping review. Rev Enferm UFSM 2021 Jul 22;11:e56. [doi: 10.5902/2179769264393]
9. do Carmo Barbosa AL, de São Joaquim BS. The practice of therapeutic guidance in the daily life of the rehabilitation professional. Escola de Educação Física, Fisioterapia e Terapia Ocupacional. URL: http://www.eeffto.ufmg.br/eeffto/biblioteca/monografia/560/ [accessed 2024-04-29]
10. Peters MD, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Scoping reviews. In: Aromataris E, Munn Z, editors. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; Jul 27, 2016:513-551.
11. 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 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
12. Orientações terapêuticas para autogestão do Transtorno Depressivo Maior: revisão de escopo. Open Science Framework. URL: https://osf.io/z3ya7/ [accessed 2024-04-29]
13. Aromataris E, Lockwood C, Porritt K, Pilla B, Jordan Z. JBI Manual for Evidence Synthesis. Adelaide, Australia: Joanna Briggs Institute; 2024.
14. Geoffroy PA, Hoertel N, Etain B, Bellivier F, Delorme R, Limosin F, et al. Insomnia and hypersomnia in major depressive episode: prevalence, sociodemographic characteristics and psychiatric comorbidity in a population-based study. J Affect Disord 2018 Jan 15;226:132-141. [doi: 10.1016/j.jad.2017.09.032] [Medline: 28972930]
15. Leggett AN, Conroy DA, Blow FC, Kales HC. Bright light as a preventive intervention for depression in late-life: a pilot study on feasibility, acceptability, and symptom improvement. Am J Geriatr Psychiatry 2018 May;26(5):598-602 [FREE Full text] [doi: 10.1016/j.jagp.2017.11.007] [Medline: 29269208]
16. Luca A, Luca M, Calandra C. Sleep disorders and depression: brief review of the literature, case report, and nonpharmacologic interventions for depression. Clin Interv Aging 2013;8:1033-1039 [FREE Full text] [doi: 10.2147/CIA.S47230] [Medline: 24019746]
17. Ma Y, Yeung A, Yang AC, Peng C, Clain A, Alpert J, et al. The effects of Tai Chi on sleep quality in Chinese American patients with major depressive disorder: a pilot study. Behav Sleep Med 2018 Sep 27;16(4):398-411. [doi: 10.1080/15402002.2016.1228643] [Medline: 27676270]
18. Xu G, Xiao Q, Huang B, Lei H, Yin Z, Huang L, et al. Clinical evidence for association of acupuncture with improved major depressive disorder: a systematic review and meta-analysis of randomized control trials. Neuropsychobiology 2023 Oct 21;82(1):1-13. [doi: 10.1159/000527903] [Medline: 36549277]
19. Farrar AJ, Farrar FC. Clinical aromatherapy. Nurs Clin North Am 2020 Dec;55(4):489-504 [FREE Full text] [doi: 10.1016/j.cnur.2020.06.015] [Medline: 33131627]
20. Ter Heege FM, Mijnster T, van Veen MM, Pijnenborg GH, de Jong PJ, Boersma GJ, et al. The clinical relevance of early identification and treatment of sleep disorders in mental health care: protocol of a randomized control trial. BMC Psychiatry 2020 Jun 24;20(1):331 [FREE Full text] [doi: 10.1186/s12888-020-02737-3] [Medline: 32580724]
21. Rethorst CD, Greer TL, Toups MS, Bernstein I, Carmody TJ, Trivedi MH. IL-1β and BDNF are associated with improvement in hypersomnia but not insomnia following exercise in major depressive disorder. Transl Psychiatry 2015 Aug 04;5(8):e611 [FREE Full text] [doi: 10.1038/tp.2015.104] [Medline: 26241349]
22. Kuyken W, Warren FC, Taylor RS, Whalley B, Crane C, Bondolfi G, et al. Efficacy of mindfulness-based cognitive therapy in prevention of depressive relapse: an individual patient data meta-analysis from randomized trials. JAMA Psychiatry 2016 Jun 01;73(6):565-574 [FREE Full text] [doi: 10.1001/jamapsychiatry.2016.0076] [Medline: 27119968]

XSL•FO
RenderX

23. Chung K, Kim S, Lee E, Park JY. Mobile app use for insomnia self-management in urban community-dwelling older Korean adults: retrospective intervention study. JMIR Mhealth Uhealth 2020 Aug 24;8(8):e17755 [FREE Full text] [doi: 10.2196/17755] [Medline: 32831177]
24. de Sousa Ibiapina AR, Lopes-Junior LC, Veloso LU, Costa AP, da Silva Júnior FJ, e Silva Sales JC, et al. Effects of music therapy on anxiety and depression symptoms in adults diagnosed with mental disorders: a systematic review. Acta Paul Enferm 2022;35:e2212. [doi: 10.37689/acta-ape/2022AR0]
25. Prathikanti S, Rivera R, Cochran A, Tungol JG, Fayazmanesh N, Weinmann E. Treating major depression with yoga: a prospective, randomized, controlled pilot trial. PLoS One 2017 Mar 16;12(3):e0173869 [FREE Full text] [doi: 10.1371/journal.pone.0173869] [Medline: 28301561]
26. Yang L. Chinese medicine for major depressive disorder: clinical evidence, patients' experience and expectations, and doctors' perceptions. Beijing University of Chinese Medicine. 2019. URL: https://tinyurl.com/39hrasy4 [accessed 2024-07-29]
27. Russi MA, Sperandio RD, De Oliveira CR, Sanches LR. Physical therapy intervention in the treatment of insomnia. Braz J Dev 2022 Nov 03;8(11):70997-71015. [doi: 10.34117/bjdv8n11-017]
28. Bellingham GL. Spiritually focused mindfulness meditation: an interpretative phenomenological analysis of the effect of spiritually focused mindfulness meditation on depression with a clinical population. Liverty University. 2019. URL: https://tinyurl.com/8hvh29n9 [accessed 2023-04-09]
29. Hulme K, Safari R, Thomas S, Mercer T, White C, Van der Linden M, et al. Fatigue interventions in long term, physical health conditions: a scoping review of systematic reviews. PLoS One 2018 Oct 12;13(10):e0203367 [FREE Full text] [doi: 10.1371/journal.pone.0203367] [Medline: 30312325]
30. Macnamara CL, Cvejic E, Parker GB, Lloyd AR, Lee G, Beilharz JE, et al. Personalised relaxation practice to improve sleep and functioning in patients with chronic fatigue syndrome and depression: study protocol for a randomised controlled trial. Trials 2018 Jul 11;19(1):371 [FREE Full text] [doi: 10.1186/s13063-018-2763-8] [Medline: 29996933]
31. Ito M, Miyamae M, Yokoyama C, Yamashita Y, Ueno O, Maruo K, et al. Augmentation of positive valence system-focused cognitive behavioral therapy by inaudible high-frequency sounds for anhedonia: a trial protocol for a pilot study. JAMA Netw Open 2019 Nov 01;2(11):e1915819 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.15819] [Medline: 31747035]
32. Braun JT, Al Shirawi MI, Rotzinger S, Kennedy SH, Bartel L. A pilot study investigating the effect of music-based intervention on depression and anhedonia. Front Psychol 2019 May 8;10:1038 [FREE Full text] [doi: 10.3389/fpsyg.2019.01038] [Medline: 31133945]
33. Ebrahem SM, Masry SE. Effect of relaxation therapy on depression, anxiety, stress and quality of life among diabetic patients. Clin Nurs Stud 2017 Jan 19;5(1):35. [doi: 10.5430/cns.v5n1p35]
34. Kinser PA, Bourguignon C, Taylor AG, Steeves R. "A feeling of connectedness": perspectives on a gentle yoga intervention for women with major depression. Issues Ment Health Nurs 2013 Jun 28;34(6):402-411 [FREE Full text] [doi: 10.3109/01612840.2012.762959] [Medline: 23805925]
35. Grandia PM. Patient-initiated strategies for self-management of depression and low mood: understanding theory and changing behaviour. Thesis University of Ottawa. 2015. URL: https://ruor.uottawa.ca/items/f5a6ccef-9bfd-4eee-bddb-d3418bda5e9a [accessed 2023-02-22]
36. Sathyanarayanan G, Vengadavaradan A, Bharadwaj B. Role of yoga and mindfulness in severe mental illnesses: a narrative review. Int J Yoga 2019;12(1):3-28 [FREE Full text] [doi: 10.4103/ijoy.IJOY_65_17] [Medline: 30692780]
37. van Grieken RA, Kirkenier AC, Koeter MW, Nabitz UW, Schene AH. Patients' perspective on self-management in the recovery from depression. Health Expect 2015 Oct 02;18(5):1339-1348 [FREE Full text] [doi: 10.1111/hex.12112] [Medline: 23910797]
38. Sarsak HI. Applied occupational therapy for major depressive disorder: clinical case report. MOJ Clin Med Case Rep 2020;10(2):31-34. [doi: 10.15406/mojcr.2020.10.00337]
39. Souza CD. Chinese auricular acupuncture in the treatment of depression. University of São Paulo. 2019. URL: https://www.teses.usp.br/teses/disponiveis/22/22131/tde-31012020-111959/pt-br.php [accessed 2024-04-30]
40. Birgitta GA, Wagman P, Hedin K, Håkansson C. Treatment of depression and/or anxiety - outcomes of a randomised controlled trial of the tree theme method® versus regular occupational therapy. BMC Psychol 2018 May 23;6(1):25 [FREE Full text] [doi: 10.1186/s40359-018-0237-0] [Medline: 29792226]
41. Duggal HS. Self-management of depression: beyond the medical model. Perm J 2019;23:55 [FREE Full text] [doi: 10.7812/TPP/18-295] [Medline: 31167709]
42. Naslund JA, Whiteman KL, McHugo GJ, Aschbrenner KA, Marsch LA, Bartels SJ. Lifestyle interventions for weight loss among overweight and obese adults with serious mental illness: a systematic review and meta-analysis. Gen Hosp Psychiatry 2017 Jul;47:83-102 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.04.003] [Medline: 28807143]
43. de Cassia Rondina R, Aragão Martins R, da Silva AL, Mascarenhas GP, Travenssolo JG, Luiz dos Santos B. Practicing physical exercise and symptoms of depression in college students. Mundo Saúde 2018 Jul 30;42(3):710-727 [FREE Full text] [doi: 10.15343/0104-7809.20184203710727]

XSL•FO
RenderX

44. Daumit GL, Dickerson FB, Wang NY, Dalcin A, Jerome GJ, Anderson CA, et al. A behavioral weight-loss intervention in persons with serious mental illness. N Engl J Med 2013 Apr 25;368(17):1594-1602 [FREE Full text] [doi: 10.1056/NEJMoa1214530] [Medline: 23517118]
45. Kazemi A, Noorbala AA, Azam K, Eskandari MH, Djafarian K. Effect of probiotic and prebiotic vs placebo on psychological outcomes in patients with major depressive disorder: a randomized clinical trial. Clin Nutr 2019 Apr;38(2):522-528. [doi: 10.1016/j.clnu.2018.04.010] [Medline: 29731182]
46. Colombari FS, Stafford RS, Heaney CA, Rosas LG. The effect of a behavioral weight-loss intervention on depressive symptoms among Latino immigrants in a randomized controlled trial. J Immigr Minor Health 2018 Oct 16;20(5):1182-1189 [FREE Full text] [doi: 10.1007/s10903-017-0663-8] [Medline: 29038966]
47. Carraça EV, Silva MN, Coutinho SR, Vieira PN, Minderico CS, Sardinha LB, et al. The association between physical activity and eating self-regulation in overweight and obese women. Obes Facts 2013;6(6):493-506 [FREE Full text] [doi: 10.1159/000356449] [Medline: 24217426]
48. de Vaz Pato Oom MC. Physical activity in the treatment of young adult depressive disease. University of Beira Interior. URL: http://hdl.handle.net/10400.6/10748 [accessed 2023-04-14]
49. Kandola A, Ashdown-Franks G, Hendrikse J, Sabiston CM, Stubbs B. Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity. Neurosci Biobehav Rev 2019 Dec;107:525-539. [doi: 10.1016/j.neubiorev.2019.09.040] [Medline: 31586447]
50. Barton J, Griffin M, Pretty J. Exercise-, nature- and socially interactive-based initiatives improve mood and self-esteem in the clinical population. Perspect Public Health 2012 Mar 07;132(2):89-96. [doi: 10.1177/1757913910393862] [Medline: 22616429]
51. Soares TD, Peroza LR, Cerezer M, Nedel SS, Branco JC. Effects of physical exercise on obesity and depression: a review. Braz J Obes Nutr Weigh Loss 2020;14(86):511-518 [FREE Full text] [Medline: 1981-9919]
52. Brush CJ, Hajcak G, Bocchine AJ, Ude AA, Muniz KM, Foti D, et al. A randomized trial of aerobic exercise for major depression: examining neural indicators of reward and cognitive control as predictors and treatment targets. Psychol Med 2022 Apr;52(5):893-903. [doi: 10.1017/S0033291720002573] [Medline: 32838817]
53. Belvederi MM, Ekkekakis P, Magagnoli M, Zampogna D, Cattedra S, Capobianco L, et al. Physical exercise in major depression: reducing the mortality gap while improving clinical outcomes. Front Psychiatry 2018 Jan 10;9:762 [FREE Full text] [doi: 10.3389/fpsyt.2018.00762] [Medline: 30687141]
54. Turner AP, Hartoonian N, Hughes AJ, Arewasikporn A, Alschuler KN, Sloan AP, et al. Physical activity and depression in MS: the mediating role of behavioral activation. Disabil Health J 2019 Oct;12(4):635-640. [doi: 10.1016/j.dhjo.2019.04.004] [Medline: 31097413]
55. Toups M, Carmody T, Greer T, Rethorst C, Grannemann B, Trivedi MH. Exercise is an effective treatment for positive valence symptoms in major depression. J Affect Disord 2017 Feb;209:188-194 [FREE Full text] [doi: 10.1016/j.jad.2016.08.058] [Medline: 27936452]
56. Archer T, Josefsson T, Lindwall M. Effects of physical exercise on depressive symptoms and biomarkers in depression. CNS Neurol Disord Drug Targets 2014 Jan 22;13(10):1640-1653. [doi: 10.2174/1871527313666141130203245] [Medline: 25470398]
57. Bains N, Abdijadid S. Major depressive disorder. StatPearls. URL: https://www.ncbi.nlm.nih.gov/books/NBK559078/ [accessed 2024-06-23]
58. Bajaj B, Robins RW, Pande N. Mediating role of self-esteem on the relationship between mindfulness, anxiety, and depression. Pers Individ Differ 2016 Jul;96:127-131. [doi: 10.1016/j.paid.2016.02.085]
59. Randal C, Pratt D, Bucci S. Mindfulness and self-esteem: a systematic review. Mindfulness 2015 Apr 9;6(6):1366-1378. [doi: 10.1007/S12671-015-0407-6]
60. Farb N, Anderson A, Ravindran A, Hawley L, Irving J, Mancuso E, et al. Prevention of relapse/recurrence in major depressive disorder with either mindfulness-based cognitive therapy or cognitive therapy. J Consult Clin Psychol 2018 Feb;86(2):200-204. [doi: 10.1037/ccp0000266] [Medline: 29265831]
61. Katterman SN, Kleinman BM, Hood MM, Nackers LM, Corsica JA. Mindfulness meditation as an intervention for binge eating, emotional eating, and weight loss: a systematic review. Eat Behav 2014 Apr;15(2):197-204. [doi: 10.1016/j.eatbeh.2014.01.005] [Medline: 24854804]
62. Mindfulness based cognitive therapy and emotion focused therapy as treatments for major depression. Gena Fawns. 2013. URL: https://genafawns.com.au/wp-content/uploads/2016/01/Mindfulness_and_EFT_for_Depression_2013.pdf [accessed 2023-02-01]
63. Economides M, Martman J, Bell MJ, Sanderson B. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness (N Y) 2018;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
64. Person SM. A two-study investigation into the link between rumination and night eating, and symptom improvement following a mindfulness-based intervention. Ontario Lakehead University. URL: https://knowledgecommons.lakeheadu.ca/handle/2453/4741 [accessed 2023-04-09]

XSL•FO
RenderX

65. Schanche E, Vøllestad J, Visted E, Svendsen JL, Osnes B, Binder PE, et al. The effects of mindfulness-based cognitive therapy on risk and protective factors of depressive relapse - a randomized wait-list controlled trial. BMC Psychol 2020 Jun 05;8(1):57 [FREE Full text] [doi: 10.1186/s40359-020-00417-1] [Medline: 32503649]
66. Rech DL, Schmidt KE, Rudnicki T, Schmidt MM. Techniques for managing the emotion of anger: a systematic review. Stud Res Psychol 2022 Apr 25;22(1):292-307 [FREE Full text] [doi: 10.12957/epp.2022.66485]
67. Moloud R, Saeed Y, Mahmonir H, Rasool GA. Cognitive-behavioral group therapy in major depressive disorder with focus on self-esteem and optimism: an interventional study. BMC Psychiatry 2022 Apr 28;22(1):299 [FREE Full text] [doi: 10.1186/s12888-022-03918-y] [Medline: 35484528]
68. Dobkin RD, Mann SL, Interian A, Gara MA, Menza M. Cognitive behavioral therapy improves diverse profiles of depressive symptoms in Parkinson's disease. Int J Geriatr Psychiatry 2019 May 04;34(5):722-729 [FREE Full text] [doi: 10.1002/gps.5077] [Medline: 30714202]
69. Santos CH. Efficacy of procedural cognitive therapy and behavioral activation in the treatment of Major Depressive Disorder: a randomized clinical trial. Federal University of Bahia. 2017. URL: https://sucupira-legado.capes.gov.br/sucupira/public/consultas/coleta/trabalhoConclusao/viewTrabalhoConclusao.jsf?popup=true&id_trabalho=6013054 [accessed 2023-04-10]
70. Ahern E, Kinsella S, Semkovska M. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev Pharmacoecon Outcomes Res 2018 Feb 30;18(1):25-41. [doi: 10.1080/14737167.2018.1407245] [Medline: 29145746]
71. Korrelboom K, Maarsingh M, Huijbrechts I. Competitive memory training (COMET) for treating low self-esteem in patients with depressive disorders: a randomized clinical trial. Depress Anxiety 2012 Feb;29(2):102-110. [doi: 10.1002/da.20921] [Medline: 22495940]
72. Wang X, Zhou X, Yang H. Guided self-help behavioral activation intervention for geriatric depression: protocol for pilot randomized controlled trial. JMIR Res Protoc 2020 Sep 24;9(9):e18259 [FREE Full text] [doi: 10.2196/18259] [Medline: 32518060]
73. Alves KI, Bonvicini CR. The role of behavioral activation in the management of depressive symptoms. Res Soc Develop 2022 Jan 04;11(1):e15311123711 [FREE Full text] [doi: 10.33448/rsd-v11i1.23711]
74. Wisenthal A, Krupa T, Kirsh B, Lysaght R. Insights into cognitive work hardening for return-to-work following depression: qualitative findings from an intervention study. Work 2019;62(4):599-613 [FREE Full text] [doi: 10.3233/WOR-192893] [Medline: 31104044]
75. Berman MI, Morton SN, Hegel MT. Uncontrolled pilot study of an acceptance and commitment therapy and health at every size intervention for obese, depressed women: accept yourself!. Psychotherapy (Chic) 2016 Dec;53(4):462-467 [FREE Full text] [doi: 10.1037/pst0000083] [Medline: 27797546]

Abbreviations

JBI: Joanna Briggs Institute
PCC: population, concept, and context
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

Edited by T de Azevedo Cardoso; submitted 04.07.24; peer-reviewed by MJ Norton, E Oliveira; comments to author 01.10.24; revised version received 21.10.24; accepted 19.12.24; published 06.03.25.

Please cite as:
de Campos Tibúrcio P, Maria Marcheti P, Miori Pascon D, Montebello Junior MA, Alzete de Lima M, Fernandes CS, Samarina Vilaça de Brito Santos C, Socorro de Sousa Nóbrega MDP
Therapeutic Guidelines for the Self-Management of Major Depressive Disorder: Scoping Review
Interact J Med Res 2025;14:e63959
URL: https://www.i-jmr.org/2025/1/e63959
doi:10.2196/63959
PMID:40053745

©Priscila de Campos Tibúrcio, Priscila Maria Marcheti, Daniela Miori Pascon, Marco Antônio Montebello Junior, Maria Alzete de Lima, Carla Sílvia Fernandes, Célia Samarina Vilaça de Brito Santos, Maria do Perpétuo Socorro de Sousa Nóbrega. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 06.03.2025. 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

XSL•FO
RenderX

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

Comparison of Outcomes Between Staged and Same-Day Circumferential Spinal Fusion for Adult Spinal Deformity: Systematic Review and Meta-Analysis

Mert Marcel Dagli[1], MD; Ryan William Turlip[1], BA; Felix C Oettl[2,3], MD; Mohamed Emara[4]; Jaskeerat Gujral[1]; Daksh Chauhan[1], BS; Hasan S Ahmad[1], MD; Gabrielle Santangelo[1], MD; Connor Wathen[1], MD; Yohannes Ghenbot[1], MD; John D Arena[1], MD; Joshua L Golubovsky[1], MD; Ben J Gu[1], MD; John H Shin[1], MD; Jang Won Yoon[1], MD, MSc; Ali K Ozturk[1], MD; William C Welch[1], MD

[1]Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[2]Hospital for Special Surgery, New York, NY, United States

[3]Department of Orthopedic Surgery, Balgrist University Hospital, University of Zurich, Zurich, Switzerland

[4]College of Medicine, University of Sharjah, Sharjah, United Arab Emirates

Corresponding Author:
Mert Marcel Dagli, MD
Department of Neurosurgery
Perelman School of Medicine
University of Pennsylvania
801 Spruce Street
Philadelphia, PA, 19107
United States
Phone: 1 2672306493
Email: Marcel.Dagli@Pennmedicine.upenn.edu

Abstract

Background: Adult spinal deformity (ASD) is a prevalent condition often treated with circumferential spinal fusion (CF), which can be performed as staged or same-day procedures. However, evidence guiding the choice between these approaches is lacking.

Objective: This study aims to compare patient outcomes following staged and same-day CF for ASD.

Methods: Following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines, a comprehensive literature search was conducted in PubMed, MEDLINE, Embase, Cochrane CENTRAL, Web of Science, and Scopus. Eligibility criteria included studies comparing outcomes following staged and same-day CF in adults with ASD. Searches were exported to Covidence, and records were deduplicated automatically. Title and abstract screening, full-text review, and data extraction were performed by two independent reviewers, with all conflicts being resolved by a third reviewer. A meta-analysis was conducted for outcomes reported in 3 or more studies.

Results: Seven studies with 741 patients undergoing CF for ASD were included in the review (staged: n=331, 44.7% and same-day: n=410, 55.3%). Four studies that had comparable outcomes were merged for the quantitative meta-analysis and split based on observed measures. The meta-analysis revealed significantly shorter hospital length of stay (mean difference 3.98, 95% CI 2.23-5.72 days; P<.001) for same-day CF. Three studies compared the operative time between staged and same-day CF, with all reporting a lower mean operative time for same-day CF (mean between 291-479, SD 129 minutes) compared to staged CF (mean between 426-541, SD 124 minutes); however, inconsistent reporting of mean and SD made quantitative analyses unattainable. Of the 4 studies that compared estimated blood loss (EBL) in the relevant groups, 3 presented a lower EBL (mean between 412-1127, SD 954 mL) in same-day surgery compared to staged surgery (mean between 642, SD 550 to 1351, SD 869 mL). Both studies that reported intra- and postoperative adverse events showed more intraoperative adverse events in staged CF (10.9% and 13.6%, respectively) compared to same-day CF (9.1% and 3.6%, respectively). Four studies measuring any perioperative adverse events showed a higher incidence of adverse events in staged CF than all studies combined. However, quantitative analysis of EBL, intraoperative adverse events, and perioperative adverse events found no statistically significant difference. Postoperative adverse events, reoperation, infection rates, and readmission rates showed inconsistent findings between studies. Data quality assessment revealed a moderate degree of bias for all included studies.

XSL•FO
RenderX

Conclusions: Same-day CF may offer shorter operating time and hospital stay compared to staged CF for ASD. However, there was marked heterogeneity in perioperative outcomes reporting, and continuous variables were inconsistently presented. This underscored the need for standardized reporting of clinical variables and patient-reported outcomes and higher evidence of randomized controlled trials to elucidate the clinical superiority of either approach.

Trial Registration: PROSPERO CRD42022339764; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=339764

International Registered Report Identifier (IRRID): RR2-10.2196/42331

(Interact J Med Res 2025;14:e67290) doi:10.2196/67290

KEYWORDS

adults; circumferential fusion; scoliosis; spinal curvature; spinal fusion; spinal deformity; intraoperative; postoperative; perioperative; systematic reviews; meta-analysis; PRISMA

Introduction

Adult spinal deformity (ASD) is defined as abnormal curvature of the spine and is becoming increasingly prevalent, affecting up to 68% of the older adult population [1,2]. ASD is a complex spectrum of spinal pathology, including deformities such as lordosis, kyphosis, or scoliosis of the lumbar and thoracic spinal column. Although untreated adolescent ASD does occur, it typically presents in patients older than 60 years due to factors such as age-related spinal degeneration or reduced bone density [1,3].

Individuals with ASD can undergo expectant (observation alone), nonoperative, or operative therapies. At present, there is no high-quality evidence to support the decisions surgeons and patients face in treatment selection [4]. In past years, pain management and physical therapy were the preferred treatment options for ASD due to the high risk of adverse events, prolonged recovery time, and financial burden associated with surgical intervention [1,5]. If nonsurgical approaches fail to improve patients' quality of life, surgical intervention is often considered. Multicenter retrospective cohort studies previously showed an improvement in patient-reported outcomes following the surgical management of ASD [3,6]. Indications for surgery include (1) progressive curvature of the spine with sagittal or coronal imbalance, (2) significant loss of pulmonary function caused by the misalignment and deformity, and (3) loss of function due to pain associated with spinal curvature [7-10]. These are weighted against the patient comorbidities and risks of operation [11].

Long-segment surgical management by circumferential spinal fusion (CF) has increased in popularity due to its added stability granted by both anterior and posterior fixation of the spinal column [12]. CF attempts to remedy the limitations of lateral approaches alone, such as the need for intraoperative patient repositioning, which increases operative time and puts the patient at risk for adverse events due to longer time under anesthesia [13-15]. ASD can be treated by CF in 2 primary ways: staged and same day. Staged fusions occur on 2 distinct operative days, while same-day fusions are completed within a single session. Staging is largely determined by surgeon preference and case complexity, which can cause variability in the clinical management of ASD. The preference to treat with or without staging does not necessitate a gold-standard treatment for a given case complexity but rather can depend on surgical training differences and hospital administration pressures. To our knowledge, there has not been a review of published literature on staging in CF. This systematic review and meta-analysis aims to assess and quantify the patient outcomes after staged and same-day CF for ASD to guide operative decision-making and patient selection.

Methods

Guidelines, Protocol, and Registration

The design and reporting of this study were supported by the following guidelines: PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 1) and PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) [16,17]. In accordance with PRISMA-P guidelines, the protocol of the systematic review was registered on the PROSPERO (CRD42022339764) and disseminated through *JMIR Research Protocols* (PRR1-10.2196/4233), with the protocol being published before any data were collected [12]. There were no deviations from the protocol.

Eligibility Criteria

The Population, Intervention, Comparison, Outcome (PICO) framework was used to formulate inclusion criteria.

- Population: Adults with ASD
- Intervention: Staged CF surgery
- Comparison: Same-day CF surgery
- Outcomes: Perioperative outcomes (estimated blood loss [EBL], operative time, and length of hospital stay), adverse events, infection rates, and hospital readmissions or reoperations

Studies that do not differ in surgical timing (staged vs same day), nonhuman or adolescent patient populations, reviews, conference abstracts, single-case studies, or technical notes were excluded from the analysis. Further, only studies originally published in English were considered.

Search Strategy

Databases explored included PubMed, MEDLINE, Embase, Cochrane CENTRAL, Web of Science, and Scopus. A literature search was conducted in accordance with the PRISMA guidelines on August 2, 2023. We used a complex search string that was modified and fitted to the unique search functions of each queried database (Multimedia Appendix 2). Additional

XSL•FO
RenderX

searching through gray literature and reference lists was conducted to identify studies not initially captured by the database query.

Data Selection and Extraction

Studies and full text were screened, data were extracted using Covidence (Veritas Health Innovation), and duplicates were automatically removed by the software [18]. Titles and abstracts were first screened independently by 2 reviewers (FCO and ME). Next, the full text of each paper was assessed by 2 reviewers (FCO and ME) to determine the eligibility of the studies. At both stages, a third reviewer (MMD) resolved any conflicts. The following data were extracted by two authors (FCO and ME) with a third (MMD) resolving any conflicts: authors; publication year; location; number of patients; age; study type; population details; surgery details; and results, including intraoperative adverse events, postoperative adverse events, postoperative infection, perioperative adverse events, hospital length of stay (LOS), intensive care unit (ICU) LOS, reoperation, readmission, and patient-reported outcomes.

Data Quality

The ROBINS-I (Cochrane) tool was used to assess the risk of bias in the included nonrandomized studies, covering bias due to confounding variables, patient selection, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of reported results [19]. Two reviewers (FCO and JG) independently scored all domains, with a third reviewer (MMD) resolving any conflicts. Robvis was used for figure generation [20].

Data Synthesis

Studies with comparable outcomes were merged for the quantitative meta-analysis and split based on observed measures. After conducting a qualitative evaluation, we determined that there were sufficient data to perform a meta-analysis. RevMan (version 8.4; Cochrane) using random-effects modeling was used for all quantitative analyses. Mean differences for continuous variables (surgical time, EBL, hospital LOS, and ICU LOS) and odds ratios for categorical variables (intraoperative adverse events, postoperative adverse events, postoperative infection, any adverse events, readmission, and reoperation) were the end points of the meta-analysis.

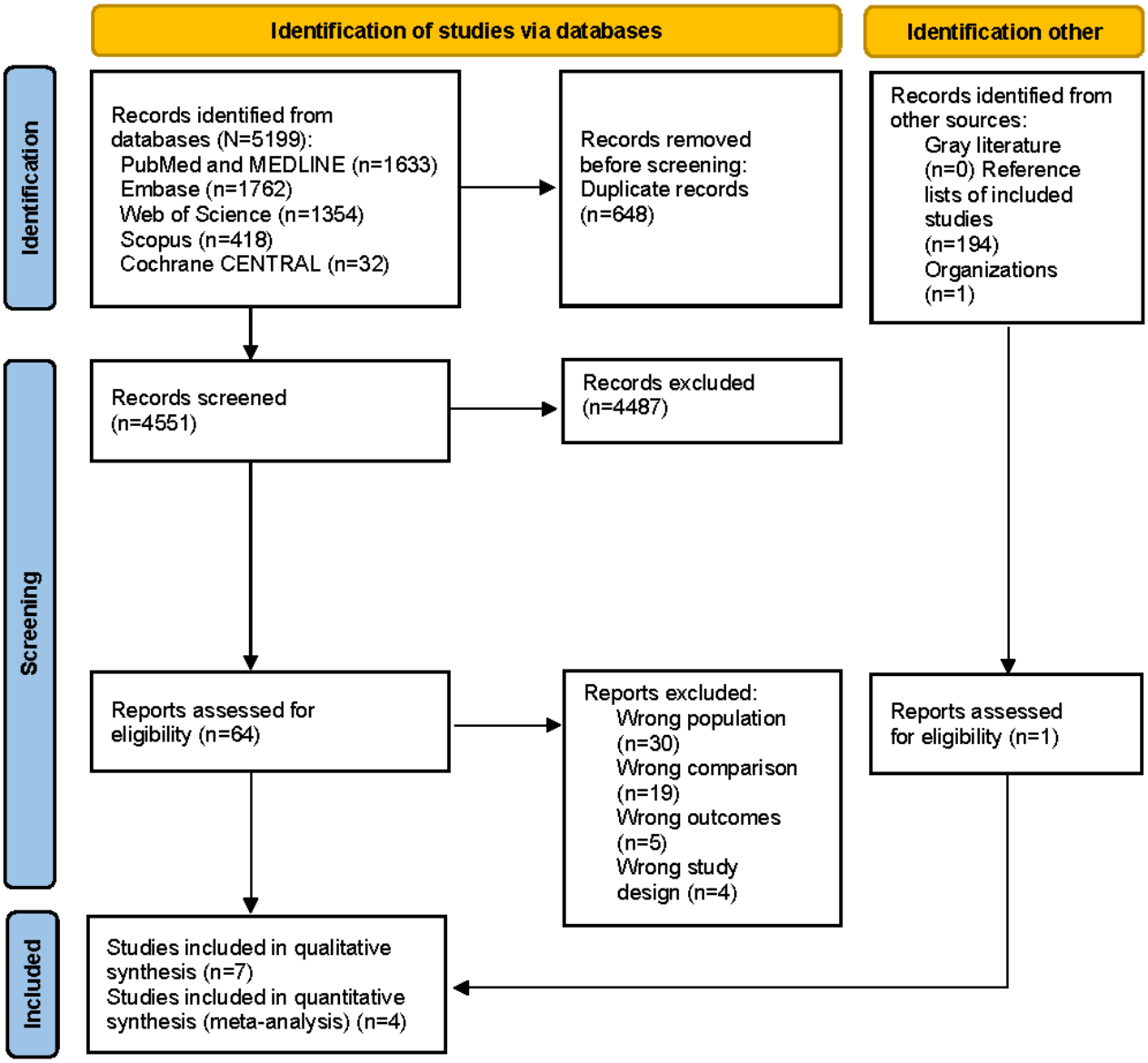
Results

Study Identification

In our search (Figure 1), we identified 5199 unique studies by searching PubMed, MEDLINE, Embase, Cochrane CENTRAL, Web of Science, and Scopus, which were included for abstract screening, of which 64 were forwarded for full-text screening. After full-text review, 7 original studies were included in the data extraction process and 4 of them were in the quantitative analysis [21-27]. Studies were excluded during the full-text review for the following reasons: wrong comparator (n=19, 29.7%), wrong patient population (n=19, 29.7%), pediatric population (n=11, 17.2%), wrong outcomes reported (n=5, 7.8%), and wrong study design (n=4, 6.3%).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analysis) flow diagram.



Baseline Parameters

Six of the studies included were conducted in the United States [21-25,27], while 1 study was conducted in Japan [26]. The included studies describe a total of 741 patients undergoing either staged (n=331, 44.7%) or same-day (n=410, 55.3%) CF for ASD: 297 (40.1%) patients underwent anterior lumbar interbody fusion, 408 (55.1%) patients underwent lateral interbody fusion (extreme lateral interbody fusion, direct lateral interbody fusion, and lateral lumbar interbody fusion), 54 (7%) patients underwent either lateral lumbar interbody fusion or transforaminal lumbar interbody fusion, and all 741 patients underwent posterior spinal fixation. The average number of fused vertebrae across studies ranged from 4.4 to 10 (SD 3.9). The largest variation between groups within a study ranged from 7.3 (SD 3.1) in the same-day group versus 10 (SD 3.9) in the staged group. The combined mean vertebrae fused for staged and same-day CF was 7.54 (SD 2.41) and 6.62 (SD 2.40), respectively. The follow-up period over all included studies ranged from 1 to more than 36 months, and the average age of patients ranged from 58.8 (SD 9.0) to 72.3 years. The patients included in the study of Masuda et al [26] and Albayar et al [21] were inverse probability weighted to facilitate the comparison of spinal deformity and control for any differences between the groups (Table 1).

XSL•FO
RenderX

Table 1. Patient demographics and significant results of included studies.

Authors (year)	Country	Patients	Age	Study type	Population details and differences	Surgery details	Results
Albayar et al (2023) [21]	United States	• n=44 • n=56	• Mean (SD): 58.8 (9.0) years • Mean (SD): 62.0 (11.9) years	• Retrospective cohort study, inverse probability weighted	• Patients aged >18 years at the time of surgery and diagnoses of ASDa undergoing (ALIFb), and open posterior lumbar or thoracolumbar (PSFc)	• Staged ALIF, and open posterior lumbar or thoracolumbar PSF, post. Vertebrae fused: mean 10 (SD 3.9) • Same-day ALIF, and open posterior lumbar or thoracolumbar PSF, post. Vertebrae fused: mean 7.3 (SD 3.1)	• Staged: EBLd mean 1351.7 (SD 869) mL, LOSe mean 10.5 (SD 5) days, IOAEf (n=6), POAEg (n=30), reoperation (n=10), POIh (n=5), and readmission (n=10). • Same day: EBL mean 1127.6 (SD 945.4), LOS mean 6.2 (SD 3.1) days, IOAE (n=2), POAE (n=30), reoperation (n=8), POI (n=1), and readmission (n=8).
Anand et al (2014) [22]	United States	• n=37 • n=13	• Mean (range): 61 (20-85) years	• Retrospective cohort study	• Patients with adult idiopathic scoliosis corrections undergoing cMISSi, Cobb angle of greater than 30 but less than 75 degrees	• Staged DLIFj and L5-S1 XLIFk followed by PSF. Vertebrae fused: mean 7 (range 4-15) • Same-day DLIF and L5-S1 XLIF followed by PSF. Vertebrae fused: mean 7 (range 4-15)	• Staged: EBL mean 763 (range 25-2500) mL and ORl time mean 482 (range 83-546) minutes. • Same day: EBL mean 613 (range 150-1500) ml and OR time mean 351 (range 176-510) minutes.
Anand et al (2013) [23]	United States	• n=36 • n=35	• Mean (range): 64 (20-84) years	• Retrospective cohort study	• Adults with scoliosis undergoing cMISS, two or more levels	• Stage combination of DLIF and XLIF with PSF. Vertebrae fused: mean 4.4 • Same-day combination of DLIF and XLIF with PSF. Vertebrae fused: mean 4.4	• Staged: EBL 671 mL and OR time 426 minutes. • Same day: EBL 412 mL and OR time 291 minutes.
Arzeno et al (2019) [24]	United States	• n=45 • n=47	• Mean (95% CI): 68 (61-78) years • Mean (95% CI): 68 (62-72) years	• Retrospective cohort study	• Patients with ASD, undergoing anterior (including lateral and anterolateral approaches) and PSF of at least 5 levels • Groups differ in approach, Ponte osteotomy, 3-column osteotomy, O-arm, neuromonitoring, decompression, number of posterior levels, fused; number. of osteotomy levels, mean; and number of decompression levels	• Staged circumferential spinal fusion (anterior, posterior), Ponte osteotomy (n=39), 3-column osteotomy (n=7), and decompression (n=34). Vertebrae fused: mean 8 (95% CI 5-9) • Same-day circumferential spinal fusion (anterior, posterior), Ponte osteotomy (n=24), 3-column osteotomy (n=1), and decompression (n=16). Vertebrae fused: mean 9 (95% CI 9-9)	• Staged: LOS mean 9 days, reoperation (n=5), readmission (n=1), POI (n=2), and PEAEm (n=12). • Same day: LOS mean 6 days, reoperation (n=7), readmission (n=6), POI (n=3), and PEAE (n=7).

XSL•FO
RenderX

Authors (year)	Country	Patients	Age	Study type	Population details and differences	Surgery details	Results
Harris et al (2021) [25]	United States	• n=41 • n=46	• Mean (SD): 62.0 (8.3) years • Mean (SD): 60.0 (13.0) years	• Retrospective cohort study	• Patients with ASD who underwent long PSF (more than 5 levels fused, with fusion to the pelvis) • Groups differ in previous spine surgery, scoliosis or kyphosis, pseudarthrosis, and pelvic incidence	• Staged circumferential -ALIF and PSF. Vertebrae fused: mean 8.7 (SD 0.48) • Same-day circumferential – ALIF and PSF. Vertebrae fused: mean 7.4 (SD 2.4)	• Staged: ODIn mean 45 (SD 17) and SRS-22ro mean 2.8 (SD 0.6) • Same day: ODI mean 48 (SD 15) and SRS-22r mean 2.8 (SD 0.6)
Masuda et al (2023) [26]	Japan	• n=101 • n=186	• Mean (SD): 72.2 (7.3) years • Mean (SD): 72.4 (8.0) years	• Retrospective cohort study, propensity score weighted	• Patients with ASD, ≥4 fused levels and at least 1 level using LLIFp, and presence of at least 1 spinal deformity marker: scoliosis Cobb angle≥20°, sagittal vertical axis≥5 cm, pelvic tilt≥25°, pelvic incidence minus lumbar lordosis angle≥10°, and thoracic kyphosis≥60°	• Staged circumferential-LLIF and PSF. Vertebrae fused: mean 7.7 (SD 2.3) • Same day circumferential-LLIF and PSF. Vertebrae fused: mean 6.2 (SD 2.4)	• Staged: EBL mean 642.5 (SD 550.5) mL, OR time mean 541.3 (SD 124.1) minutes, LOS mean 42 (SD 25) days, IOAE (n=11), POAE (n=11), reoperation (n=11), POI (n=4), and AAE[q] (n=22). • Same day: EBL 722.2 (SD 612.6) mL, OR time mean 479.9 (SD 128.5) minutes, LOS mean 34.1 (SD 18.2) days, IOAE (n=17), POAE (n=23), reoperation (n=19), POI (n=5), and AAE (n=40).
Than et al (2019) [27]	United States	• n=27 • n=27	• N/Ar • N/A	• Retrospective cohort study	• Patients with ASD, coronal Cobb angle >20, SVAs >5 cm, pelvic tilt >20, pelvic incidence–LLt >10, and thoracic kyphosis >60.	• Staged MISu LLIF or MIS TLIFv with PSF. Vertebrae fused: mean 5.4 • Same-day MIS LLIF or MIS TLIF with PSF. Vertebrae fused: mean 5.3	• Staged: Reoperation (n=4), POI (n=0), and PEAE (n=9). • Staged: Reoperation (n=7), readmission (n=1), POI (n=1), and PEAE (n=8).

[a]ASD: adult spinal deformity.

[b]ALIF: anterior lumbar interbody fusion.

[c]PSF: posterior spinal fusion.

[d]EBL: estimated blood loss.

[e]LOS: length of stay.

[f]IOAE: intraoperative adverse events

[g]POAE: postoperative adverse events.

[h]POI: postoperative infection.

[i]cMISS: circumferential minimally invasive spinal surgery.

[j]DLIF: direct lateral interbody fusion.

[k]XLIF: extreme lateral interbody fusion.

[l]OR: operating room.

[m]PEAE: perioperative adverse events.

[n]ODI: Oswestry Disability Index.

[o]SRS-22r: Scoliosis Research Society-22 revised.

XSL•FO
RenderX

[p]LLIF: lateral lumbar interbody fusion.

[q]AAE: any adverse events.

[r]N/A: not available.

[s]SVA: sagittal vertical axis.

[t]LL: lumbar lordosis.

[u]MIS: minimally invasive surgery.

[v]TLIF: transforaminal lumbar interbody fusion.

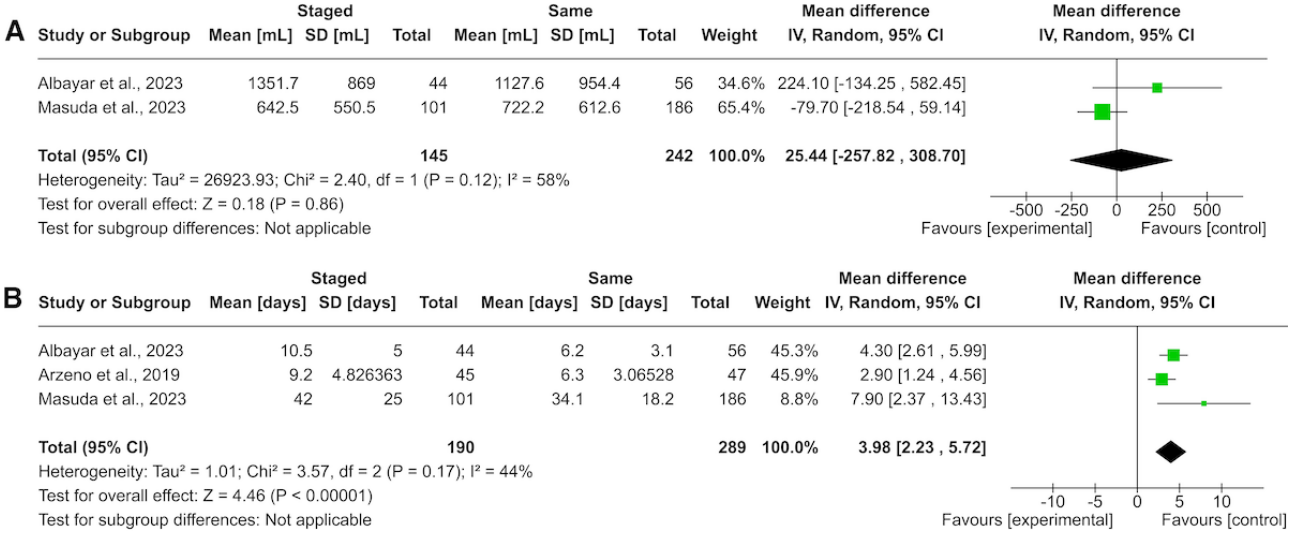
Quantitative Analysis

Of the included studies, 4 studies compared EBL in the relevant groups, with 3 studies presenting a lower EBL (mean between 412-1127, SD 954 mL) in same-day surgery compared to staged surgery (mean between 642, SD 550, and 1351, SD 869 mL) [21-24]. The meta-analysis shows a nonsignificant advantage for same-day surgery **(**Figure 2A [21,24,26]). Only 2 studies that measured EBL were included in the quantitative analysis because of inconsistencies in reporting, where some did not report variables as measures of variance, which made pooling in these instances not feasible. Three studies compared the operative time between staged and same-day CF, with all of them reporting a lower mean operative time for same-day CF (mean between 291-479 minutes) compared to staged CF (mean between 426-541 minutes) [22-24]. Just 1 group reported mean and SD for odds ratio time, thus restricting the potential for a quantitative analysis.

Three studies comparatively evaluated the hospital LOS [21,22,25]. All three studies consistently found that the mean LOS was less for same-day CF (mean between 6-34.1 days) in comparison to staged CF (mean between 9-42 days). The meta-analysis clearly presented a shorter LOS in patients undergoing same-day CF compared to staged CF (Figure 2B).

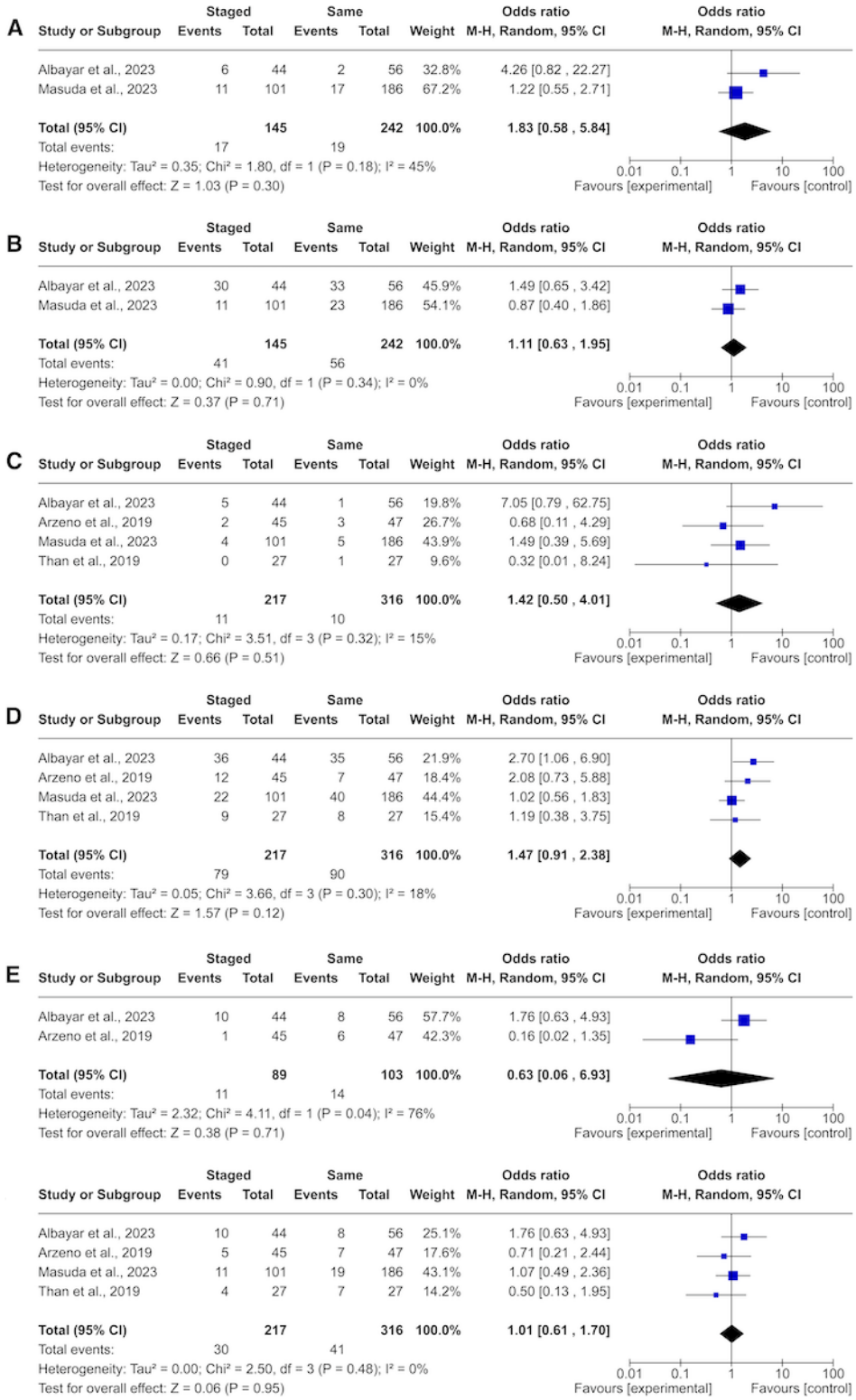
Two of the 7 studies compared intraoperative and postoperative adverse events between staged and same-day CF procedures [21,26]. Both Masuda et al [26] and Albayar et al [21] reported more intraoperative adverse events in staged CF (10.9% and 13.6%, respectively) compared to same-day CF (9.1% and 3.6%, respectively); however, the meta-analysis failed to show a statistically significant difference between the groups (Figure 3A [21-27]). Masuda et al [26] reported fewer postoperative adverse events in staged CF (10.9 vs 12.4%), while Albayar et al [21] presented a lower incidence in same-day CF (53.6 vs 68.2%), without a significant difference in the meta-analysis (Figure 3B). Four studies measured any perioperative adverse events [21,22,26,27]. The overarching analysis showed a higher incidence of adverse events in Staged CF over all studies; however, the meta-analysis did not show significance (Figure 3D).

Figure 2. Forest plots comparing (A) estimated blood loss and (B) hospital length of stay between patients who underwent staged or same-day circumferential spinal fusion for adult spinal deformity. IV: inverse variance.



XSL•FO
RenderX

Figure 3. Forest plots comparing (A) intraoperative adverse event, (B) postoperative adverse event, (C) postoperative infection, (D) perioperative adverse event, (E) 30-day readmission, and (F) reoperation rates between patients who underwent staged or same-day circumferential spinal fusion for adult spinal deformity. M-H: Mantel-Haenszel.



Four of the included studies compared reoperation and postoperative infection rates between patient groups [21,22,26,27]. Albayar et al [21] and Masuda et al [26] showed a slightly lower postoperative infection in same-day CF versus staged CF (4.4 vs 6.4%, 2.6 vs 4%, and 1.7 vs 11.4% respectively), while Than et al [27] showed a lower postoperative infection in staged CF (0% vs 3.7%;Figure 3C). While Arzeno et al [24] and Than et al [27] reported a lower reoperation in patients undergoing staged CF (11.1 vs 14.9% and 14.8 vs 25.9%, respectively). Masuda et al [26] and Albayar et al [21] reported less reoperation in patients undergoing same-day CF versus staged CF (10.2 vs 10.9% and 14.2 vs 22.7% respectively). However, none of these differences in

XSL•FO
RenderX

reoperation reached statistical significance in either the original respective studies or our meta-analysis (Figure 3F).

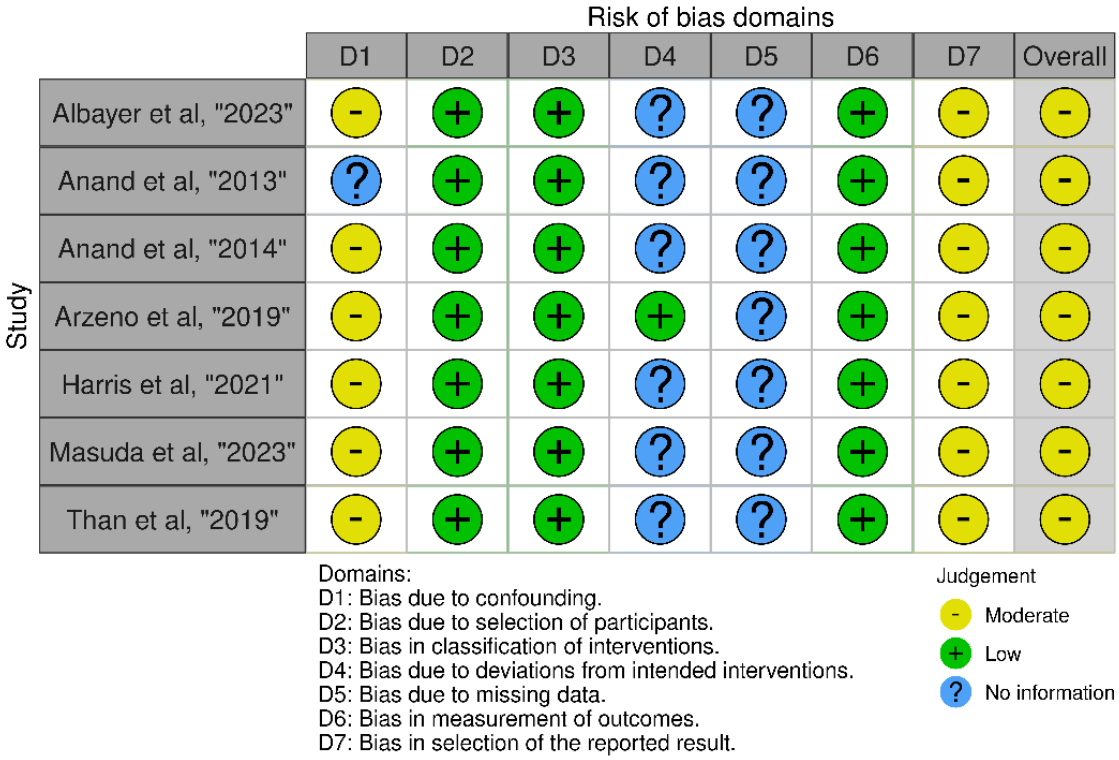
Readmission rates reported by Arzeno et al [24] and Albayar et al [21] demonstrated diverging results. Arzeno et al [24] found a lower readmission for staged CF versus same-day CF (2.2 vs 12.8%), while Albayar et al [21] reported a higher readmission in staged CF versus same-day CF (22.7 vs 14.3%). The meta-analysis did not indicate a conclusive result (Figure 3E).

Harris et al [25] are the only authors reporting on either the Oswestry Disability Index (ODI) or the Scoliosis Research Society Score, showing a better outcome measured by the ODI in same-day CF and no difference in Scoliosis Research Society Score.

Risk of Bias Analysis

There was a moderate degree of bias for all included studies (Figure 4 [21-27]).

Figure 4. ROBINS-I traffic light plot of included studies.



Discussion

Overview

In this systematic review and meta-analysis, we compared differences in perioperative outcome variables for individuals undergoing staged or same-day CF in 7 included studies. The meta-analysis revealed a statistically significant increase in hospital LOS in the staged group when compared to the same-day group. There were no statistically significant differences in EBL, intraoperative adverse event, postoperative adverse event, postoperative infection, perioperative adverse event, 30-day readmission, and reoperation rates between patients who underwent staged or same-day CF for ASD. There were marked differences between patient populations and the subsequent clinical outcomes that were reported in each observational study, which most did not adjust for. Our quantitative results are paradoxical as there were inconsistencies between studies in the staged and same-day subgroups for certain reported outcomes, while other findings had stronger conclusive evidence. Due to the apparent differences in consensus, the research included in the following discussion is organized according to the consistency of results.

Mixed Findings

Heterogeneity in meta-analyses grants the opportunity to examine variable factors that may be leading to the results. There were several reported outcomes that yielded inconclusive evidence in this study. Adverse event rates were reported by authors in several ways, with some splitting between intraoperative or postoperative adverse events, while others reported any adverse event throughout the perioperative course. Two studies reported more intraoperative adverse events in staged CF, but the meta-analysis failed to show a statistically significant difference. This can indicate that there is an increased operative risk when staging, but it can also be due to the generally sicker patients with more spinal deformity that is leading to increased intraoperative adverse events. Further, 1 study presented a lower postoperative adverse events rate during staged CF, while a different group found a higher rate in staged, indicating no difference in postoperative complications between the 2 approaches. When all perioperative adverse events were instead considered, every study reported a higher incidence in staged CF, and this result trended toward significance. Vastly different conclusions depending on the stage of surgery in which the adverse events are considered illuminate the need for greater consistency in reporting outcomes. However, the increased perioperative adverse events seen in staged CF, although not

XSL•FO
RenderX

statistically significant, could be clinically important in decision-making.

The Centers for Medicare & Medicaid Services targeted 30-day readmission rates as a source of unnecessary costs [28]. A study conducted by McCarthy et al [29] found the total hospital costs to surgically treat ASD averaged US $120,394, with primary surgery averaging US $103,143, and total readmission costs of US $67,262 for their cohort. The high costs associated with spine surgery make patient readmission and reoperation lucrative targets. There is profound heterogeneity in reoperation and readmission rates that depend on several factors, such as patient demographics, procedure types, and institutional factors. A recent systematic review found the 30-day readmission rate in spine surgery to be 4.2% and 7.4% [30]. In this review, 2 studies reported readmission with diverging results. Arzeno et al [24] found a lower readmission for staged CF, while Albayar et al [21] reported a higher readmission for same-day CF. Likewise, 4 studies reported reoperation with a split in consensus [21,24,26,27]. Given the small number of studies that measured reoperation and readmission, the results from our meta-analysis were inconclusive. The lack of statistically significant differences in the quantitative analysis combined with the heterogeneity in the qualitative review indicates that no differences in readmission and reoperation rates exist between the 2 groups.

It was also uncommon for the included studies to report specific reasons that reoperation or readmission occurred. As readmission depends on many variables including patient comorbidities, initial risk, or operative adverse events, it is crucial for authors to include these measures in future comparative studies for stronger subgroup analyses. Postoperative infection is also a driver leading to reoperation or readmission, as well as additional health care costs [31]. A recent meta-analysis found the pooled incidence of surgical site infection in 22,475 patients to be 3.1% [32]. Our included studies also reported a low incidence of postoperative infection; however, there were inconsistencies as 3 out of the 4 groups found a lower postoperative infection in same-day CF [21,24,26,27]. There are several risk factors for postoperative infection that may lead to heterogeneity. Farshad et al [33] found that intraoperative EBL was a risk factor for postoperative infection. Some studies found BMIs to be predictors of postoperative infection, although this result remains controversial [34-36]. The discrepancies seen in this review may stem from variations in surgical techniques, perioperative antibiotic prophylaxis, institutional differences, and patient-specific risk factors. The results of this study's meta-analysis do not support any difference between the 2 techniques in postoperation infection risk.

Associations of Operative Variables in CF Staging

For highly invasive spine surgeries, such as long-segmented CF for ASD, patients may require extended resource use [37]. Surgical operative time and LOS are critical metrics in evaluating the efficiency and resource use of staging in CF. Shorter surgical times are generally associated with reduced intraoperative adverse events, lower anesthesia-related risks, and decreased blood loss, thus contributing to improved patient safety and outcomes [15,38]. Peng et al [39] conducted a meta-analysis on correlations between operative time and postoperative infection, and they concluded that there was a 4-fold increase in postoperative infection risk for operations greater than 3 hours [39]. The studies included in our review all found a lower mean operative time for same-day CF.

A shorter LOS is also desirable as it minimizes health care resource use and lowers costs [40]. Further, early mobilization and discharge are associated with improved patient satisfaction and reduced psychological stress [41]. However, reducing LOS has also been associated with an increasing readmission rate at the population level [42]. Our meta-analysis found a significantly lower LOS for same-day CF compared to staged CF. This difference in LOS may be attributed to the cumulative effect of multiple hospital admissions, prolonged recovery periods between surgeries, and the need for additional preoperative preparation in staged CF. While the lower operative time and LOS results in patients undergoing same-day CF seem intuitive, it is important to interpret them cautiously, considering potential factors such as patient selection, discharge criteria, and institutional practices that may differ between studies. Moreover, while shorter operative time and LOS may reduce health care costs and improve resource allocation, they should not compromise patient safety or postoperative care quality.

The results of shorter LOS and operative time combined with the lack of conclusive differences between groups for adverse events, readmission, reoperation, and infection rates indicate that patient prognosis is similar when between staging preferences. The shorter time spent in the hospital and operating room is clinically important and contributes to improved patient safety and outcomes while reducing hospital resource use. However, the potential advantages of staging may outweigh these benefits and thus improve patient prognosis for more complex cases with a higher degree of relative deformity. Therefore, the results from this review and meta-analysis show preference toward same-day CF, while staging may largely depend on a patient's condition. Further studies investigating the impact of patient-specific factors, for example, varying degrees of deformity severity, on treatment selection and subsequent outcomes are required.

Limitations

There are several limitations to this study and the papers included in this review. First, the limited literature comparing staged and same-day CF resulted in only 7 studies being included, which reduced the statistical power of the meta-analysis as only 4 studies qualified for quantitative analysis. All the included studies are retrospective cohort studies with a level of evidence of III or II, and only 2 studies used inverse probability weighting to control for between-patient differences. Consequently, there was a moderate potential for bias, which limited our ability to reach generalizable conclusions. To elucidate the differences in surgical treatment options for CF, it will be necessary to investigate staging in level I or II randomized controlled trials (RCTs). Without RCTs, it is difficult to adjust for surgeon preference bias and the complexity of the case. Each study also demonstrated heterogeneity with respect to the patient outcomes reported,

XSL•FO
RenderX

further limiting the ability for robust statistical analysis. Continuous variables were generally poorly reported since many did not report measures of variance such as mean and SD, but rather mean or median and range (min-max or IQR). This made pooling in those instances not feasible, limiting the generalizability of results. Further, only 1 group presented patient-reported outcomes, such as ODI, which is a standardized clinical variable that can be useful for measuring subjective pain and disability. Novel studies should aim to investigate patient-reported outcomes and standardize the variables being reported, so future meta-analyses will have wider samples to yield more conclusive evidence for staging differences in CF. In doing so, the mixed results presented in this study can be further assessed to support wider implementation of either staged or same-day CF while integrating clinically significant metrics to evaluate patient outcomes.

Conclusions

Here, we present the first systematic review and quantitative meta-analysis on staging in CF to treat patients with ASD. Operative time and hospital LOS were significantly lower in same-day CF surgery, with EBL and perioperative adverse events also trending toward significance. However, there was heterogeneity in additional operative measures, such as intra- or postoperative adverse event rates, reoperation, and readmission, and no differences were found in our meta-analysis. Based on our results, it is suggested that same-day CF is advantageous as a potential time-save with patients spending less time in the operating room and hospital, potentially saving costs. However, it is still unclear whether either same-day or staged CF provides a clinical advantage for patient outcomes. Additional level I and II RCTs should be conducted to elucidate the associations between these variables and provide stronger evidence in favor of either approach.

Acknowledgments

This work was partially supported by the Catherine Sharpe Foundation.

Data Availability

All data generated or analyzed during this study are included in this published paper.

Authors' Contributions

WCW is the guarantor of the study. MMD and WCW led the conceptualization, data acquisition, analysis, and drafting and revision of the manuscript. RWT, JG, GS, DC, HSA, CAW, TG, JDA, JLG, and BJG contributed to data acquisition, analysis, and drafting. Independent review was performed by FCO, ME, and MMD. All authors contributed to the analysis, interpretation, and drafting. JHS, JWY, AKO, and WCW contributed critical guidance at all stages of the study. The manuscript was reviewed and edited, and its final version was approved by all authors.

Conflicts of Interest

In the past 36 months, AKO has received consulting fees from Medacta and Johnson and Johnson and has served as an E2M ad hoc reviewer for the Journal of Neurosurgery Publishing Group (JNS PG), with no relation to this work. Additionally, within the same period, JWY has received a grant from Johnson and Johnson; consulting fees from Medyssey, TrackX, Richard Wolf, and Johnson and Johnson; holds patents planned, issued, or pending with Kinesiometrics and MedCyclops; and has served in a leadership role on the Scientific Program Committee of the American Association of Neurological Surgeons (AANS)/Congress of Neurological Surgeons (CNS) Joint Section on Disorders of the Spine and Peripheral Nerves, with no relation to this work. All other authors report no conflict of interest.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 86 KB - ijmr_v14i1e67290_app1.pdf]

Multimedia Appendix 2
Supplemental search strings for query.
[DOCX File , 15 KB - ijmr_v14i1e67290_app2.docx]

References

1. Kim HJ, Yang JH, Chang D, Lenke LG, Suh SW, Nam Y, et al. Adult spinal deformity: a comprehensive review of current advances and future directions. Asian Spine J 2022;16(5):776-788 [FREE Full text] [doi: 10.31616/asj.2022.0376] [Medline: 36274246]
2. Cerpa M, Lenke LG, Fehlings MG, Shaffrey CI, Cheung KMC, Carreon LY. Evolution and advancement of adult spinal deformity research and clinical care: an overview of the Scoli-RISK-1 study. Global Spine J 2019;9:8S-14S [FREE Full text] [doi: 10.1177/2192568219828729] [Medline: 31157151]

XSL•FO
RenderX

3. Diebo BG, Shah NV, Boachie-Adjei O, Zhu F, Rothenfluh DA, Paulino CB, et al. Adult spinal deformity. Lancet 2019;394(10193):160-172. [doi: 10.1016/S0140-6736(19)31125-0] [Medline: 31305254]
4. Youssef JA, Orndorff DO, Patty CA, Scott MA, Price HL, Hamlin LF, et al. Current status of adult spinal deformity. Global Spine J 2013;3(1):51-62 [FREE Full text] [doi: 10.1055/s-0032-1326950] [Medline: 24436852]
5. Aebi M. The adult scoliosis. Eur Spine J 2005;14(10):925-948. [doi: 10.1007/s00586-005-1053-9] [Medline: 16328223]
6. Kim HJ, Chang D. Clinical and radiographic parameters for patients with adult spinal deformity. J Korean Med Assoc 2021;64(11):743-747. [doi: 10.5124/jkma.2021.64.11.743]
7. Bradford DS, Tay BK, Hu SS. Adult scoliosis: surgical indications, operative management, complications, and outcomes. Spine 1999;24(24):2617-2629. [doi: 10.1097/00007632-199912150-00009] [Medline: 10635525]
8. Glassman SD, Berven S, Kostuik J, Dimar JR, Horton WC, Bridwell K. Nonsurgical resource utilization in adult spinal deformity. Spine 2006;31(8):941-947. [doi: 10.1097/01.brs.0000209318.32148.8b] [Medline: 16622386]
9. Acaroglu E, Yavuz AC, Guler UO, Yuksel S, Yavuz Y, Domingo-Sabat M, et al. A decision analysis to identify the ideal treatment for adult spinal deformity: Is surgery better than non-surgical treatment in improving health-related quality of life and decreasing the disease burden? Eur Spine J 2016;25(8):2390-2400. [doi: 10.1007/s00586-016-4413-8] [Medline: 26821143]
10. Ames CP, Scheer JK, Lafage V, Smith JS, Bess S, Berven SH, et al. Adult spinal deformity: epidemiology, health impact, evaluation, and management. Spine Deform 2016;4(4):310-322. [doi: 10.1016/j.jspd.2015.12.009] [Medline: 27927522]
11. Good CR, Auerbach JD, O'Leary PT, Schuler TC. Adult spine deformity. Curr Rev Musculoskelet Med 2011;4(4):159-167 [FREE Full text] [doi: 10.1007/s12178-011-9101-z] [Medline: 22021017]
12. Dagli MM, Narang S, Malhotra K, Santangelo G, Wathen C, Ghenbot Y, et al. The differences between same-day and staged (circumferential) fusion surgery in adult spinal deformity: protocol for a systematic review. JMIR Res Protoc 2022;11(11):e42331 [FREE Full text] [doi: 10.2196/42331] [Medline: 36441570]
13. Thomas JA, Menezes C, Buckland AJ, Khajavi K, Ashayeri K, Braly BA, et al. Single-position circumferential lumbar spinal fusion: an overview of terminology, concepts, rationale and the current evidence base. Eur Spine J 2022;31(9):2167-2174. [doi: 10.1007/s00586-022-07229-4] [Medline: 35913621]
14. Kim BD, Hsu WK, de Oliveira GS, Saha S, Kim JYS. Operative duration as an independent risk factor for postoperative complications in single-level lumbar fusion: an analysis of 4588 surgical cases. Spine 2014;39(6):510-520. [doi: 10.1097/BRS.0000000000000163] [Medline: 24365901]
15. Hersey AE, Durand WM, Eltorai AEM, DePasse JM, Daniels AH. Longer operative time in elderly patients undergoing posterior lumbar fusion is independently associated with increased complication rate. Global Spine J 2019;9(2):179-184 [FREE Full text] [doi: 10.1177/2192568218789117] [Medline: 30984498]
16. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
17. 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;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
18. Covidence Systematic Review Software, Veritas Health Innovation. URL: https://www.covidence.org/ [accessed 2025-01-15]
19. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
20. McGuinness LA, Higgins JPT. Risk-of-bias VISualization (robvis): An R package and shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
21. Albayar A, Santangelo G, Spadola M, Macaluso D, Ali ZS, Saifi C, et al. Comparison of staged vs same-day circumferential spinal fusions for adult spinal deformity. Int J Spine Surg 2023;17(6):843-855 [FREE Full text] [doi: 10.14444/8548] [Medline: 37827708]
22. Anand N, Baron EM, Khandehroo B. Is circumferential minimally invasive surgery effective in the treatment of moderate adult idiopathic scoliosis? Clin Orthop Relat Res 2014;472(6):1762-1768 [FREE Full text] [doi: 10.1007/s11999-014-3565-2] [Medline: 24658900]
23. Anand N, Baron EM, Khandehroo B, Kahwaty S. Long-term 2- to 5-year clinical and functional outcomes of minimally invasive surgery for adult scoliosis. Spine 2013;38(18):1566-1575. [doi: 10.1097/BRS.0b013e31829cb67a] [Medline: 23715025]
24. Arzeno AH, Koltsov J, Alamin TF, Cheng I, Wood KB, Hu SS. Short-term outcomes of staged versus same-day surgery for adult spinal deformity correction. Spine Deform 2019;7(5):796-803.e1. [doi: 10.1016/j.jspd.2018.12.008] [Medline: 31495481]
25. Harris AB, Puvanesarajah V, Raad M, Marrache M, Ren M, Skolasky RL, et al. How is staging of ALIF following posterior spinal arthrodesis to the pelvis related to functional improvement in patients with adult spinal deformity? Spine Deform 2021;9(4):1085-1091. [doi: 10.1007/s43390-020-00272-5] [Medline: 33464551]

XSL•FO
RenderX

26. Masuda S, Fujibayashi S, Takemoto M, Ota M, Onishi E, Odate S, et al. Association of two-staged surgery with systemic perioperative complications in lateral lumbar interbody fusion for adult spinal deformity: a propensity score-weighted study. Eur Spine J 2023;32(3):950-956. [doi: 10.1007/s00586-023-07539-1] [Medline: 36680618]
27. Than KD, Park P, Tran S, Mundis GM, Fu K, Uribe JS, et al. Analysis of complications with staged surgery for less invasive treatment of adult spinal deformity. World Neurosurg 2019;126:e1337-e1342. [doi: 10.1016/j.wneu.2019.03.090] [Medline: 30898739]
28. McIlvennan CK, Eapen ZJ, Allen LA. Hospital readmissions reduction program. Circulation 2015;131(20):1796-1803 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.010270] [Medline: 25986448]
29. McCarthy IM, Hostin RA, Ames CP, Kim HJ, Smith JS, Boachie-Adjei O, et al. Total hospital costs of surgical treatment for adult spinal deformity: an extended follow-up study. Spine J 2014;14(10):2326-2333. [doi: 10.1016/j.spinee.2014.01.032] [Medline: 24469004]
30. Bernatz JT, Anderson PA. Thirty-day readmission rates in spine surgery: systematic review and meta-analysis. Neurosurg Focus 2015;39(4):E7. [doi: 10.3171/2015.7.FOCUS1534] [Medline: 26424347]
31. Anderson PA, Savage JW, Vaccaro AR, Radcliff K, Arnold PM, Lawrence BD, et al. Prevention of surgical site infection in spine surgery. Neurosurgery 2017;80(3S):S114-S123. [doi: 10.1093/neuros/nyw066] [Medline: 28350942]
32. Zhou J, Wang R, Huo X, Xiong W, Kang L, Xue Y. Incidence of surgical site infection after spine surgery: a systematic review and meta-analysis. Spine 2020;45(3):208-216. [doi: 10.1097/BRS.0000000000003218] [Medline: 31464972]
33. Farshad M, Bauer DE, Wechsler C, Gerber C, Aichmair A. Risk factors for perioperative morbidity in spine surgeries of different complexities: a multivariate analysis of 1,009 consecutive patients. Spine J 2018;18(9):1625-1631 [FREE Full text] [doi: 10.1016/j.spinee.2018.02.003] [Medline: 29452285]
34. Abdallah DY, Jadaan MM, McCabe JP. Body mass index and risk of surgical site infection following spine surgery: a meta-analysis. Eur Spine J 2013;22(12):2800-2809 [FREE Full text] [doi: 10.1007/s00586-013-2890-6] [Medline: 23828507]
35. Blam OG, Vaccaro AR, Vanichkachorn JS, Albert TJ, Hilibrand AS, Minnich JM, et al. Risk factors for surgical site infection in the patient with spinal injury. Spine 2003;28(13):1475-1480. [doi: 10.1097/01.BRS.0000067109.23914.0A] [Medline: 12838110]
36. Capen DA, Calderone RR, Green A. Perioperative risk factors for wound infections after lower back fusions. Orthop Clin North Am 1996;27(1):83-86. [doi: 10.1016/s0030-5898(20)32052-6]
37. Reis RC, de Oliveira MF, Rotta JM, Botelho RV. Risk of complications in spine surgery: a prospective study. Open Orthop J 2015;9:20-25 [FREE Full text] [doi: 10.2174/1874325001509010020] [Medline: 25674185]
38. Findlay MC, Kim RB, Warner WS, Sherrod BA, Park S, Mazur MD, et al. Identification of an operative time threshold for substantially increased postoperative complications among elderly spine surgery patients. Global Spine J 2024;14(5):1532-1541 [FREE Full text] [doi: 10.1177/21925682221149390] [Medline: 36623932]
39. Peng XQ, Sun CG, Fei ZG, Zhou QJ. Risk factors for surgical site infection after spinal surgery: a systematic review and meta-analysis based on twenty-seven studies. World Neurosurg 2019;123:e318-e329. [doi: 10.1016/j.wneu.2018.11.158] [Medline: 30496932]
40. Stephens BF, Khan I, Chotai S, Sivaganesan A, Devin CJ. Drivers of cost in adult thoracolumbar spine deformity surgery. World Neurosurg 2018;118:e206-e211. [doi: 10.1016/j.wneu.2018.06.155] [Medline: 29966783]
41. Smith G, Chirieleison S, Levin J, Atli K, Winkelman R, Tanenbaum JE, et al. Impact of length of stay on HCAHPS scores following lumbar spine surgery. J Neurosurg Spine 2019;31(3):366-371. [doi: 10.3171/2019.3.SPINE181180] [Medline: 31151093]
42. Ansari SF, Yan H, Zou J, Worth RM, Barbaro NM. Hospital length of stay and readmission rate for neurosurgical patients. Neurosurgery 2018;82(2):173-181. [doi: 10.1093/neuros/nyx160] [Medline: 28402465]

Abbreviations

ASD: adult spinal deformity
CF: circumferential spinal fusion
EBL: estimated blood loss
ICU: intensive care unit
LOS: length of stay
ODI: Oswestry Disability Index
PICO: Population, Intervention, Comparison, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols
RCT: randomized controlled trial

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 07.10.24; peer-reviewed by J Shen; comments to author 05.11.24; revised version received 31.12.24; accepted 07.01.25; published 06.03.25.

Please cite as:
Dagli MM, Turlip RW, Oettl FC, Emara M, Gujral J, Chauhan D, Ahmad HS, Santangelo G, Wathen C, Ghenbot Y, Arena JD, Golubovsky JL, Gu BJ, Shin JH, Yoon JW, Ozturk AK, Welch WC
Comparison of Outcomes Between Staged and Same-Day Circumferential Spinal Fusion for Adult Spinal Deformity: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e67290
URL: https://www.i-jmr.org/2025/1/e67290
doi:10.2196/67290
PMID:40053742

©Mert Marcel Dagli, Ryan William Turlip, Felix C Oettl, Mohamed Emara, Jaskeerat Gujral, Daksh Chauhan, Hasan S Ahmad, Gabrielle Santangelo, Connor Wathen, Yohannes Ghenbot, John D Arena, Joshua L Golubovsky, Ben J Gu, John H Shin, Jang Won Yoon, Ali K Ozturk, William C Welch. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 06.03.2025. 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

Informatics Interventions for Maternal Morbidity: Scoping Review

Jill Inderstrodt[1,2,3], MPH, PhD; Julia C Stumpff[4], MSLIS; Rebecca C Smollen[5], MPH; Shreya Sridhar[6], MPH; Sarah A El-Azab[7], MS; Opeyemi Ojo[8], MPH; Brendan Bowns[9], MPH; David A Haggstrom[3,10,11], MAS, MD

[1]Department of Health Policy and Management, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States
[2]Center for Biomedical Informatics, Regenstrief Institute, Indianapolis, IN, United States
[3]Center for Health Information and Communication, Richard L. Roudebush VA Medical Center, Indianapolis, IN, United States
[4]Ruth Lilly Medical Library, Indiana University School of Medicine, Indianapolis, IN, United States
[5]Department of Public Health, College of Health and Human Sciences, Purdue University, West Lafayette, IN, United States
[6]School of Pharmacy, University of Pittsburgh, Pittsburgh, PA, United States
[7]Department of Health Policy and Management, University of Michigan, Ann Arbor, MI, United States
[8]Department of Epidemiology, Richard M. Fairbanks School of Public Health, Indiana University, Indianapolis, IN, United States
[9]Department of Prevention, Corktown Health, Detroit, MI, United States
[10]Center for Health Services Research, Regenstrief Institute, Indianapolis, IN, United States
[11]Division of General Internal Medicine and Geriatrics, Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, United States

Corresponding Author:
Jill Inderstrodt, MPH, PhD
Department of Health Policy and Management
Richard M. Fairbanks School of Public Health
Indiana University
1050 Wishard Blvd.
Indianapolis, IN, 46202
United States
Phone: 1 3176903316
Email: ji3@iu.edu

Abstract

Background: Women have been entering pregnancy less healthy than previous generations, placing them at increased risk for pregnancy complications. One approach to ensuring effective monitoring and treatment of at-risk women is designing technology-based interventions that prevent maternal morbidities and treat perinatal conditions.

Objective: This scoping review evaluates what informatics interventions have been designed and tested to prevent and treat maternal morbidity.

Methods: MEDLINE, Embase, and Cochrane Library were searched to identify relevant studies. The inclusion criteria were studies that tested a medical or clinical informatics intervention; enrolled adult women; and addressed preeclampsia, gestational diabetes mellitus (GDM), preterm birth, Centers for Disease Control and Prevention–defined severe maternal morbidity, or perinatal mental health conditions. Demographic, population, and intervention data were extracted to characterize the technologies, conditions, and populations addressed.

Results: A total of 80 studies were identified that met the inclusion criteria. Many of the studies tested for multiple conditions. Of these, 73% (60/82) of the technologies were tested for either GDM or perinatal mental health conditions, and 15% (12/82) were tested for preeclampsia. For technologies, 32% (28/87) of the technologies tested were smartphone or tablet applications, 26% (23/87) were telehealth interventions, and 14% (12/87) were remote monitoring technologies. Of the many outcomes measured by the studies, almost half (69/140, 49%) were patient physical or mental health outcomes.

Conclusions: Per this scoping review, most informatics interventions address three conditions: GDM, preeclampsia, and mental health. There may be opportunities to treat other potentially lethal conditions like postpartum hemorrhage using proven technologies such as mobile apps. Ample gaps in the literature exist concerning the use of informatics technologies aimed at maternal morbidity. There may be opportunities to use informatics for lesser-targeted conditions and populations.

(Interact J Med Res 2025;14:e64826) doi:10.2196/64826

XSL•FO
RenderX

KEYWORDS

scoping review; maternal morbidity; medical informatics; clinical informatics; mother; pregnant; perinatal; GDM; preeclampsia; maternity; gestational diabetes mellitus

Introduction

Women have been entering pregnancy less healthy than previous generations [1]. Maternal morbidities such as preeclampsia and gestational diabetes, and severe maternal morbidities such as postpartum hemorrhage all have implications for the long-term health of mothers well beyond the postpartum period. Preeclampsia is associated with cardiovascular disease for decades following pregnancy, including chronic hypertension, stroke, and age-adjusted overall mortality [2]. Gestational diabetes mellitus (GDM), a condition that in many cases resolves itself after delivery, can still affect mothers beyond the puerperium [3]. Severe maternal morbidity (SMM) encompasses myriad conditions, including acute myocardial infarction, eclampsia, and hemorrhage. Those who have experienced SMM are more likely to die at any point after delivery and into the decades beyond the postpartum period [4]. Preterm birth (PTB; delivering before 37 weeks gestation) is associated with long-term cardiovascular complications in the mother, including ischemic heart disease, stroke, and atherosclerosis [5].

Informatics interventions offer tools that can help to prevent perinatal health conditions that have long-term health consequences for mothers, monitor these conditions in the perinatal period so that mother and baby remain healthy, and follow mothers post partum to ensure that they continue to receive the health monitoring they need. Technology-based tools that can be used for these purposes include mobile apps, wearable technology, physician decision support, and telehealth, among others. As artificial intelligence (AI) technology becomes more common in health care, a range of interventions can be used to predict, diagnose, and treat maternal morbidities that lead to maternal complications, with the hope of reducing maternal morbidity and mortality [6]. While some patients are hesitant for their physicians to use AI to diagnose or treat medical problems, physicians have demonstrated growing acceptance and belief that AI will improve interactions with their patients [7]. As the landscape of health informatics shifts substantially with the advent of new technologies, it is important to assess what technologies have already been tested to prevent, diagnose, and treat common maternal morbidities so that gaps can be identified and addressed in future research [8].

In this scoping review, we seek to identify what informatics interventions have been designed and tested that address maternal morbidity. In addition, we aim to assess the extent, range, and nature of informatics interventions that have been tested to prevent, diagnose, and treat maternal morbidities known to have long-term health consequences for mothers. The review summarizes discrete populations addressed in the literature, types of informatics interventions tested in the literature, conditions addressed by these interventions, and outcomes measured.

Methods

Search Strategy

Our scoping review methodology was guided by Arksey and O'Malley's [9] framework, stages 1-5, and Joanna Briggs's *Manual for Scoping Reviews*. We asked what informatics interventions have been designed and tested for SMM. Using the PCC (Population, Concept, and Context) framework outlined in the *Manual for Scoping Reviews*, the population was adult women with SMM as defined by the Centers for Disease Control and Prevention (CDC), the concept was informatics interventions, and the context was prospective studies anywhere in the world. We developed a review protocol to guide the process. The literature searches were led by an expert information specialist (JCS) in consultation with the research team. Four electronic databases were searched from inception until June 6, 2022: MEDLINE and Embase (searched simultaneously on Ovid), Cochrane Library (Wiley), and Engineering Village (Elsevier). One database was searched from inception until June 7, 2022: IEEE Xplore (IEEE). The MEDLINE (Ovid) search was peer-reviewed by another expert librarian using the PRESS (Peer Review of Electronic Search Strategies) checklist and modified as required [10]. The search strategy was limited to the English language and encompassed all years of publication. Embase and MEDLINE citations were deduplicated in Ovid before exporting to Covidence online review software. The full search strategies for each database are publicly available in searchRxiv [11]. Lastly, review results are reported using PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) [12]. This checklist can be found in Multimedia Appendix 1. Our protocol was registered retrospectively in Open Science Framework Registries [13].

Study Selection

The inclusion and exclusion criteria are described in Textbox 1. Studies were included that prospectively tested a medical informatics intervention on either a group of women or physicians who treat women. Study methodologies could include both quantitative and qualitative approaches. Only studies analyzing primary data to describe a health-related outcome were included. Studies were included that tested medical or clinical informatics interventions on human subjects to prevent, monitor, or treat maternal conditions that have long-term health consequences for mothers within and beyond the perinatal period. Conditions included preeclampsia or hypertensive disorders of pregnancy, GDM, PTB, new mental health diagnoses, and SMM as defined by the US CDC [14].

XSL•FO
RenderX

Textbox 1. Criteria for inclusion and exclusion.

Inclusion criteria

- Test a medical informatics intervention prospectively
- Medical or clinical informatics interventions
- Adult women or physicians who treat women
- Preeclampsia, gestational diabetes mellitus, preterm birth, perinatal mental health diagnoses, severe maternal morbidity as defined by the Centers for Disease Control and Prevention (21 indicators)
- Physical health, patient-centered outcomes, mental health, health behavior, health knowledge or attitudes, health care use, quality of care outcomes

Exclusion criteria

- Systematic reviews, scoping reviews, literature reviews, opinion pieces, commentaries, proposals, reports, conference papers
- Pilot/feasibility studies
- Bioinformatics studies, including scans, ultrasounds, biomarkers, and predictive algorithms
- Adolescents
- Preexisting conditions: type 1 or 2 diabetes, chronic hypertension, preexisting mental health conditions
- Feasibility, acceptability, user experience

Systematic reviews, scoping reviews, and literature reviews were excluded together with opinion pieces, commentaries, proposals, reports, or gray literature. Studies were also excluded if they focused primarily on bioinformatics, effectively eliminating imaging (eg, ultrasounds), biomarkers, and predictive algorithm development.

Article references were entered into review software for screening and data extraction. The first and senior authors conducted an initial title, abstract, and keyword screening. If any author recommended inclusion, then the article underwent full-text screening. All authors except for the information specialist (JCS) then conducted a full-text screening, with two team members reviewing all articles. Studies with two votes to include were moved forward for extraction; those with two votes to exclude were excluded from the study. During the full-text screening, Cohen κ was used in the initial training to gauge agreement until reviewers reached a κ of 0.80. After training, disagreements were resolved by the first author.

Data Extraction

A subset of four authors extracted data using Covidence review software. Reviewers were given an initial set of 10 articles and met with the first author for consensus before reviewing the remaining articles. All studies were reviewed for extraction by the first author and one additional author. Most data extraction categories were developed a priori, and selection options were developed by the first and senior authors after reviewing all articles for trends. The following data categories were added after the protocol had been developed: aim of the study, years of data collection, population inclusion and exclusion criteria, and total number of participants. These categories were added to the data extraction form iteratively to provide more context to the reviewed studies.

Demographic data collected for each article by reviewers included title of article, year published, lead author surname, and country in which the study was conducted. The following characteristics of included studies were extracted: informatics technology tested, health problem, aim of study, study design, years of data collection, population description, study participation inclusion criteria, study participation exclusion criteria, total number of participants, and health outcome. Reviewers were given the option of selecting multiple answers for each study characteristic. An open-ended "other" option was available for the following study characteristics: informatics technology, health problem, study design, and outcome. Following extraction, the first and senior authors also extracted data on whether a significantly positive result was found in the studies, as well as racial diversity in participant samples for those studies conducted in the United States.

Data were synthesized by tallying the options from each category (study design, region, technology, health outcome, study population, study outcome, and results) and calculating the corresponding frequencies for each option.

Results

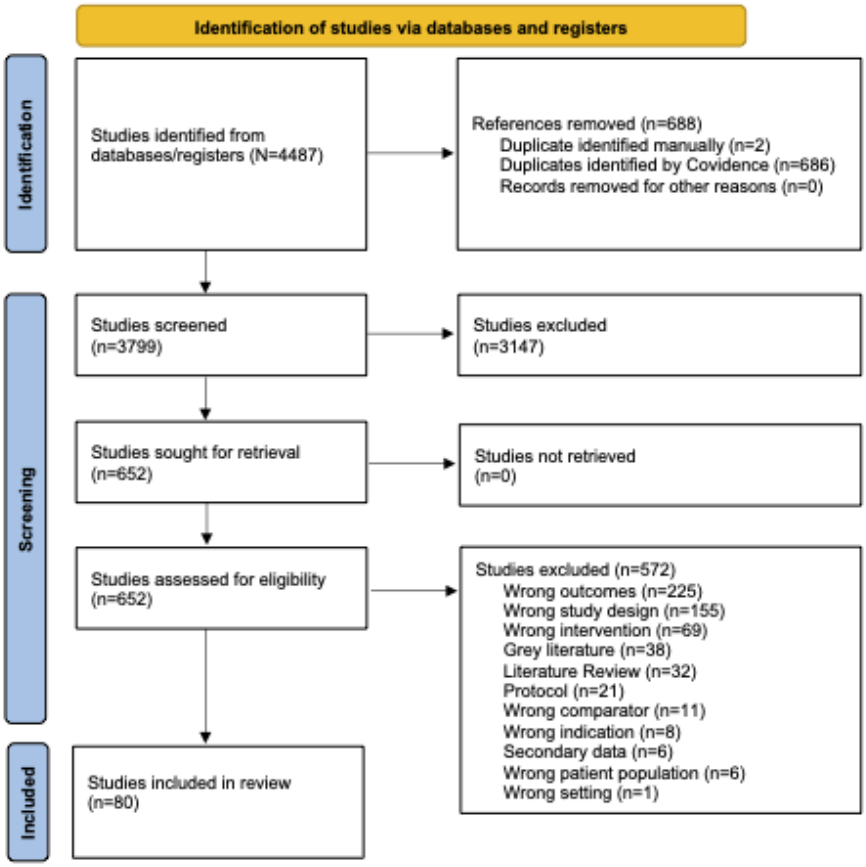
Overview

The PRISMA flow diagram for this scoping review can be found in Figure 1. A total of 3799 records were identified for title and abstract screening; 652 records underwent full-text screening for eligibility. Of these, 572 were excluded. Common reasons for exclusion were study design (eg, retrospective, systematic or other review, narrative), type of study outcome (user experience), or the wrong intervention (eg, machine learning algorithms or predictive models not tested with patients or providers). A total of 80 studies underwent full data extraction. A list of the studies and their characteristics can be found in Table 1.

Frequencies for synthesized data can be found in Table 2.

XSL•FO
RenderX

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



XSL•FO
RenderX

Table 1. Studies selected for extraction.

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Abbaspoor et al [15] 2020	Iran	• SMS text message	GDM[a]	RCT[b]	Women	100	• Medical • Health behavior
Abejirinde et al [16] 2019	Ghana	• Smartphone/tablet application • DSSc	HDP[d], GDM	Nonrandomized experimental	Women	940	• Quality of care
Al-Ofi et al [17] 2019	Saudi Arabia	• Remote monitoring • Telehealth	GDM	RCT	Women	57	• Medical • Health behavior
Arias et al [18] 2022	United States	• Telehealth	HDP	Cohort	Women	1579	• Health behavior • Health care use • Quality of care
Bartholomew et al [19] 2015	United States	• Remote monitoring • Smartphone/tablet application • Web program	GDM	RCT	Women	74	• Number of readings
Baumel et al [20] 2018	United States	• Smartphone/tablet application • Web program	Mental health condition	Nonrandomized experimental	Women	19	• Mental health
Bellad et al [21] 2020	India	• Web application	HDP	RCT	Providers	14,783	• Medical
Borgen et al [22] 2019	Norway	• Smartphone/tablet application	GDM	RCT	Women	158	• Medical • Health behavior
Carlisle et al [23] 2022	United Kingdom	• DSS	Mental health condition, PTB[e]	RCT	Women	300	• Mental health
Carroll et al [24] 2013	United States	• DSS	Mental health condition	RCT	Providers	48	• Quality of care
Chan et al [25] 2019	China	• Smartphone/tablet application	Mental health condition	RCT	Women	660	• Mental health
Chappell et al [26] 2022	United Kingdom	• Remote monitoring	HDP	RCT	Women	454	• Medical
Cheung et al [27] 2019	Australia	• SMS text messaging • Wearable	GDM	RCT	Women	60	• Medical • Number of readings • Health behavior
Dennis et al [28] 2009	Canada	• Telehealth	Mental health condition	RCT	Women	846	• Medical • Mental health
Dennis et al [29] 2020	United States	• Telehealth	Mental health condition	RCT	Women	249	• Medical • Mental health • Health care use
Felder et al [30] 2017	United States	• Web program	Mental health condition	Nonrandomized experimental	Women	27	• Medical • Mental health

XSL•FO
RenderX

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Ferrara et al [31] 2012	United States	• Telehealth	GDM	RCT	Women	1000	• Medical • Number of readings
Ferrara et al [32] 2016	United States	• Telehealth	GDM	RCT	Women	523	• Medical • Health behavior
Forsell et al [33] 2017	Sweden	• Telehealth • Web program	Mental health condition	RCT	Women	72	• Patient-centered outcomes • Mental health
Garnweidner-Holme et al [34] 2020	Norway	• Remote monitoring • Smartphone/tablet application	GDM	RCT	Women	238	• Health behavior
Ghaderi et al [35] 2019	Iran	• Smartphone/tablet application	GDM	RCT	Women	134	• Health knowledge/attitudes
Goetz et al [36] 2020	Taiwan	• Smartphone/tablet application	Mental health condition	Cohort	Women	30	• Mental health
Gong et al [37] 2021	China	• SMS text messaging	Mental health condition	RCT	Women	291	• Medical • Mental health
Guille et al [38] 2021	United States	• SMS text messaging • Telehealth	Mental health condition	Nonrandomized experimental	Women	2988	• Quality of care • Health behavior • Health care use
Guo et al [39] 2019	China	• Smartphone/tablet application	GDM	RCT	Women	172	• Medical • Number of readings • Health behavior • Health care use
Hantsoo et al [40] 2018	United States	• Smartphone/tablet application	Mental health condition	RCT	Women	61	• Health behavior • Health care use • Quality of care
Hedderson et al [41] 2018	United States	• DSS	GDM	RCT	Women	2014	• Medical
Heller et al [42] 2020	Netherlands	• Web program	Mental health condition	RCT	Women	159	• Mental health
Homko et al [43] 2007	United States	• Web program	GDM	RCT	Women	57	• Medical • Number of readings • Health knowledge/attitudes
Homko et al [44] 2012	United States	• Telehealth	GDM	RCT	Women	253	• Medical • Number of readings
Hoppe et al [45] 2020	United States	• Telehealth	HDP	Nonrandomized experimental	Women	428	• Medical • Number of readings • Health care use
	China		GDM	RCT	Women	309	• Medical

XSL•FO
RenderX

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Huang et al [46] 2021		• Smartphone/tablet application					
Jannati et al [47] 2020	Iran	• Smartphone/tablet application	Mental health condition	RCT	Women	75	• Mental health
Klokkenga et al [48] 2019	Ghana	• Smartphone/tablet application	Hemorrhage	RCT	Women	146	• Medical
Krishnamurti et al [49] 2021	United States	• Smartphone/tablet application	HDP	Cohort	Women	2567	• Health behavior • Quality of care
Lanssens et al [50] 2017	Belgium	• Remote monitoring	HDP	Retrospective	Women	166	• Medical • Health care use • Quality of care
Lanssens et al [51] 2018	Belgium	• Remote monitoring	HDP	Retrospective	Women	320	• Medical • Health care use • Quality of care
Latendresse et al [52] 2021	United States	• Telehealth	Mental health condition	Nonrandomized experimental	Women	47	• Mental health
Lewey et al [53] 2022	United States	• SMS text messaging • Web program • Wearable	HDP	RCT	Women	127	• Medical • Health behavior
Li et al [54] 2021	China	• Telehealth	GDM	RCT	Women	287	• Medical • Health behavior
Lim et al [55] 2021	Singapore	• Smartphone/tablet application	GDM	RCT	Women	200	• Medical • Mental health
Mackillop et al [56] 2018	United Kingdom	• Telehealth • SMS text messaging	GDM	RCT	Women	203	• Medical • Number of readings • Patient-centered outcomes
Milgrom et al [57] 2016	Australia	• Web program	Mental health condition	RCT	Women	43	• Mental health
Milgrom et al [58] 2021	Australia	• Telehealth	Mental health condition	RCT	Women	116	• Mental health
Miremberg et al [59] 2018	Israel	• Smartphone/tablet application	GDM	RCT	Women	120	• Medical • Number of readings • Patient-centered outcomes • Health behavior
Missler et al [60] 2020	Netherlands	• Telehealth	Mental health condition	RCT	Women	130	• Mental health • Parenting measures
Ngai et al [61] 2015	China	• Telehealth	Mental health condition	RCT	Women	397	• Mental health

XSL•FO
RenderX

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Nieminen et al [62] 2016	Sweden	• Web program	Mental health condition	RCT	Women	43	• Patient-centered outcomes • Mental health • Health behavior
Niksalehi et al [63] 2018	Iran	• SMS text messaging	Mental health condition	Nonrandomized experimental	Women	54	• Mental health
Nishimwe et al [64] 2021	Rwanda	• Smartphone/tablet application	Hemorrhage	Nonrandomized experimental	Providers	54	• Health knowledge/attitudes • Quality of care
Parsa et al [65] 2019	Iran	• Smartphone/tablet application	HDP	Nonrandomized experimental	Women	110	• Health knowledge/attitudes
Perry et al [66] 2018	United Kingdom	• Remote monitoring • Smartphone/tablet application	HDP	Case control	Women	108	• Medical • Number of readings • Health care use
Posmontier et al [67] 2016	United States	• Telehealth	Mental health condition	Cohort	Women	61	• Patient-centered outcomes • Mental health • Quality of care
Potzel et al [68] 2022	Germany	• Smartphone/tablet application	GDM	RCT	Women	203	• Medical • Number of readings • Health behavior
Rahman et al [69] 2017	Bangladesh	• Web program	Hemorrhage	RCT	Women	310	• Quality of care
Rasekaba et al [70] 2018	Australia	• Telehealth	GDM	RCT	Women	95	• Medical • Health care use
Sawyer et al [71] 2019	Australia	• Smartphone/tablet application	Mental health condition	RCT	Women	133	• Mental health
Shamshiri Milani et al [72] 2015	Iran	• Telehealth	Mental health condition	RCT	Women	126	• Mental health
Silva-Jose et al [73] 2021	Spain	• Telehealth • Web program	HDP	RCT	Women	206	• Medical • Health behavior
Simhi et al [74] 2021	Israel	• Telehealth	Mental health condition	RCT	Women	77	• Mental health • Health behavior
Skar et al [75] 2018	Norway	• Smartphone/tablet application	GDM	Qualitative	Women	17	• Mental health • Health knowledge/attitudes
St Pierre et al [76] 2017	Germany	• Anesthesia machine cognitive aid	Myocardial infarction	RCT	Women	83	• Quality of care
Sun et al [77] 2021	China	• Smartphone/tablet application	Mental health condition	RCT	Women	186	• Mental health
	Korea		GDM	RCT	Women	21	• Medical

XSL•FO
RenderX

Author (year)	Country	Technology	Condition	Study design	Population	Sample size, n	Outcome
Sung et al [78] 2019		• Remote monitoring • Smartphone/tablet application					
Sung et al [79] 2021	Korea	• Telehealth	GDM	Cohort	Women	176	• Medical • Health care use
Tian et al [80] 2020	China	• Smartphone/tablet application	GDM	RCT	Women	169	• Health behavior
Tian et al [81] 2021	China	• Telehealth	GDM	RCT	Women	309	• Medical
Triebwasser et al [82] 2020	United States	• SMS text messaging • Web-based program	HDP	RCT	Women	333	• Number of readings
Tucker et al [83] 2022	United Kingdom	• Remote monitoring • Smartphone/tablet application	HDP	RCT	Women	2346	• Medical • Mental health • Quality of care
Ugarriza and Schmidt [84] 2006	United States	• Telehealth	Mental health condition	RCT	Women	30	• Mental health
Vandenberk et al [85] 2019	Belgium	• Remote monitoring	HDP	RCT	Women	108	• Number of readings • Mental health
van den Heuvel et al [86] 2020	Netherlands	• Remote monitoring • Smartphone/tablet application	HDP	Case control	Women	133	• Medical • Health care use
van Montfort et al [87] 2020	Netherlands	• Web program	HDP	Cohort	Women	850	• Quality of care
Van Ryswyk et al [88] 2015	New Zealand	• SMS text messaging	GDM	RCT	Women	268	• Medical • Health behavior
Vigod et al [89] 2021	Canada	• Web program	Mental health condition	RCT	Women	98	• Mental health
Wozney et al [90] 2017	Canada	• Telehealth	Mental health condition	RCT	Women	62	• Mental health
Yang et al [91] 2018	China	• Smartphone/tablet application	GDM	RCT	Women	107	• Medical
Yew et al [92] 2021	Singapore	• Smartphone/tablet application	GDM	RCT	Women	340	• Medical
Zera et al [93] 2015	United States	• DSS	GDM	RCT	Women	847	• Quality of care
Zhuo et al [94] 2022	China	• Smartphone/tablet application	GDM	RCT	Women	124	• Medical • Health behavior

[a]GDM: gestational diabetes mellitus.

[b]RCT: randomized controlled trial.

[c]DSS: decision support system.

[d]HDP: hypertensive disorders of pregnancy.

[e]PTB: preterm birth.

XSL•FO
RenderX

Table 2. Outcomes measured.

	Frequency, n (%)
Study design (n=80)	
Randomized controlled trial	60 (75)
Nonrandomized experimental	9 (11)
Cohort	6 (8)
Case control	2 (3)
Retrospective	2 (3)
Qualitative	1 (1)
Region (n=80)	
North America	24 (30)
East Asia and Pacific	22 (28)
Europe and Central Asia	20 (25)
Middle East and North Africa	9 (11)
Sub-Saharan Africa	3 (4)
South Asia	2 (3)
Latin America and Caribbean	0 (0)
Technology (n=87)	
Application	28 (32)
Telehealth/telemedicine	23 (26)
Remote monitoring	12 (13)
Web-based program	10 (11)
SMS text messaging	8 (9)
EHR[a] DSS[b]	3 (3)
Wearable	2 (2)
Other	1 (1)
Perinatal health outcome (n=82)	
New mental health diagnosis	30 (37)
GDM[c]	30 (37)
Preeclampsia/HDP[d]	12 (15)
Hypertension	5 (6)
Hemorrhage	3 (4)
Preterm birth	1 (1)
Myocardial infarction	1 (1)
Study population (n=80)	
Women	77 (96)
Providers	3 (4)
Outcomes (n=140)	
Medical outcomes	39 (27.9)
Mental health outcomes	30 (21.4)
Health behavior	21 (15.0)
Quality of care	19 (13.6)
Number of readings	13 (9.3)

XSL•FO
RenderX

	Frequency, n (%)
Health care use	12 (8.6)
Health knowledge/attitudes	5 (3.6)
Other	1 (0.7)
Results (n=80)	
Significant for at least one measure	63 (79)
No significance	17 (21)

[a]EHR: electronic health record.

[b]DSS: decision support system.

[c]GDM: gestational diabetes mellitus.

[d]HDP: hypertensive disorders of pregnancy.

Included studies were mostly RCTs (n=60). Other study designs included nonrandomized experimental (n=9), cohort (n=6), case control (n=2), retrospective (n=2), and qualitative (n=1). Included studies represented a range of countries and regions. Most of the studies were conducted in North America (n=24), East Asia and the Pacific (n=22), and Western Europe (n=20). Fewer studies were authored from the Middle East and North Africa (n=9), sub-Saharan Africa (n=3), and South Asia (n=2). No studies focused on Latin America or the Caribbean.

Informatics Technologies

The specific informatics technologies tested by the study were described. Descriptive categories included telehealth or telemedicine (remote consultation via telephone or video chat that may replace medical consultation), remote monitoring (patient-performed measurements of vital signs and blood glucose), electronic health record decision support system (tools within the electronic health record that assist providers with clinical decisions), smartphone or tablet applications, SMS text messages (between providers and patients or between patients), web-based programs (patient portals and other programs administered through a web browser, not through a smartphone or tablet application), and other (with the technology charted manually). Studies tested a range of informatics technologies. The most frequently tested technologies included smartphone/tablet applications (n=28) and telehealth/telemedicine (n=23). Technologies tested less frequently included remote monitoring (n=15), web-based programs (n=10), and SMS text messaging (n=8). Technologies assessed in very few studies included electronic health record decision support (n=3), wearable technologies (n=2), and other (cognitive aid on an anesthesia machine; n=1).

Health Problems

Articles were categorized by the health problems treated by the informatics technology; some tested more than one. These included preeclampsia or hypertensive disorders of pregnancy, hypertension, GDM, PTB, hemorrhage, and new (perinatal) mental health diagnoses. Although our protocol searched for "severe maternal morbidity as defined by the CDC" and included a list of 21 conditions [13], after reviewing the articles, it was clear that the only condition from the CDC list that was frequently tested by informatics technology was postpartum hemorrhage. Thus, any additional conditions were added manually in the "other" option. Studies primarily used informatics technologies to address new mental health conditions (n=30) and GDM (n=30). Other health problems addressed by the studies include preeclampsia or hypertensive disorders of pregnancy (n=12), hypertension (n=5), and hemorrhage (n=3). One study each addressed preterm birth and myocardial infarction.

Population Type

Population type was categorized as either: women or medical providers who treat women. Included studies overwhelmingly tested informatics technologies on women, with only 3 studies testing informatics technologies on providers. US studies were also coded for sample demographics.

Outcomes

Finally, outcomes were categorized to gain an understanding of the measures upon which the studies focused; many tested more than one measure. These included medical outcomes (eg, weight loss, blood pressure, blood glucose), number of remote measures (eg, patient-collected and -reported blood pressure or blood glucose measurements), mental health outcomes (eg, Edinburgh Postnatal Depression Scale score, subjective well-being measures), health behaviors (eg, dietary habits, exercise frequency), health knowledge or attitudes (eg, patient's understanding of health condition, patient-reported attitude toward care), health care use (eg, number of emergency room visits, number of prenatal care visits), quality of care (eg, Hospital Consumer Assessment of Healthcare Providers and Systems [HCAHPS] scores, safety of care, effectiveness of care, efficiency of care), or other. Studies measured myriad outcomes, and many studies measured more than one outcome. Many assessed medical outcomes (n=39), mental health outcomes (n=30), and patient health behaviors (n=21). Fewer measured quality of care (n=19), number of remote readings (n=13), and health care use (n=12). Very few measured health knowledge and attitudes (n=5) or other outcomes (n=1).

Effect

Studies were coded as to whether a significant positive effect was exhibited for at least one health outcome measure (yes/no). Most (n=63) studies found that the informatics technology yielded at least one positive effect for a health-related (ie, not user experience) outcome. Some (n=17) studies showed no significant positive effect of the informatics technology tested.

XSL•FO
RenderX

Discussion

Principal Findings

This scoping review was performed to characterize what informatics interventions have been tested to address the maternal morbidities that result in long-term negative health outcomes for mothers. Our results show that most of the tested informatics technologies were tested in North America, East Asia, and Western Europe. They overwhelmingly focused on patient populations diagnosed with GDM, mental health conditions, and preeclampsia. Most were patient focused and most aimed to prevent, diagnose, or treat physical or mental health outcomes.

Women have been entering pregnancy less healthy than previous generations. As the world population grows increasingly obese, high BMI presents critical risks for birthing people, with obesity linked to such perinatal complications as gestational diabetes [95], preeclampsia [96], and postpartum hemorrhage [97]. Women also currently enter pregnancy in poorer cardiovascular health than previous generations, with prepregnancy hypertension especially pronounced in rural populations of childbearing age [98]. This scoping review surveys the landscape of informatics interventions tested to treat these increasingly common conditions.

Most of the technologies were tested in North America, East Asia or the Pacific, and Western Europe. Despite some evidence of resistance to new technologies [99], studies generally show perceptions of mobile health to be positive in sub-Saharan Africa, especially toward maternal health applications [100]. There may be opportunities to test technologies developed in North America, Asia, and Europe in underserved parts of the world. In addition to improving maternal health in these underserved areas, testing in diverse types of populations may yield compelling findings that could be translated into US, Asian, and European contexts through a process of reciprocal innovation [101].

Our results show that most of the tested informatics technologies focused on patient populations diagnosed with GDM; mental health conditions; and, to a lesser extent, preeclampsia. Far fewer focused on preventing or treating postpartum hemorrhage. Although prevalences of GDM and mental health conditions are high, preeclampsia and hemorrhage are leading causes of maternal mortality and pregnancy-associated deaths in both the United States and abroad [101]. Postpartum hemorrhage, in particular, is the leading cause of maternal death worldwide, causing 94% of maternal deaths [102,103].

There are at least two possible reasons for this imbalance in testing of interventions. First, diabetes mellitus and mental health conditions are also common in the general population, unlike preeclampsia and hemorrhage. Thus, preexisting interventions may be more easily transferrable to pregnant populations. Second, GDM and postpartum depression can be diagnosed through routine screening, with GDM diagnosed via prenatal glucose testing and postpartum depression diagnosed using the Edinburgh Postnatal Depression Scale. Indeed, 9.3% of interventions measured the extent/accuracy with which patients review their own biodata, including blood glucose levels. Screening for GDM could be seen as a parsimonious method to improve women's health. On the other hand, postpartum hemorrhage is less predictable, typically being diagnosed through an urgent/emergent presentation. Thus, the ease of routine screening practices may lend certain diagnoses to being more amenable to the implementation of informatics interventions.

Most of the interventions in this review aimed to prevent, diagnose, or treat physical or mental health outcomes, and almost all were patient focused. Very few interventions (n=3, 4%) were tested on providers who treat women. Of those studies, two of the interventions focused on midwives and one used a decision support system to assist physicians in screening for maternal depression. Most interventions in this review were tested on women and involved active participation from patients (eg, taking blood glucose readings, logging symptoms in an app, participating in an online support group). In other words, most of the interventions entrusted patients with aspects of their own health.

This focus on outcomes is logical given the maternal mortality crisis both within and outside of the United States. Mental health conditions are the leading cause of pregnancy-related deaths in the United States according to state Maternal Mortality Review Committees [104] and of non-Hispanic White and Hispanic women when stratified by ethnicity. Cardiac conditions are the leading cause of pregnancy-related death for Black/African American women in the United States, but the focus on interventions aimed at patients themselves minimizes the role provider knowledge and practice within a health system plays in the crisis. Research shows that lack of standardized emergency obstetric care, maternity care coordination [105], and systemic racism [106] can play a role in the safe and effective delivery of care. The results of this scoping review show a large gap in the literature related to using informatics to improve the quality of care for maternal morbidities. Interventions such as telehealth have been shown to improve obstetric quality, including perinatal smoking cessation and breastfeeding [107]. It is possible that expanding such technologies and examining the way they are used by physicians can help improve the quality of care and ultimately save the lives of women.

The need for informatics interventions does not end with birth. Prior research suggests that many physicians, including obstetricians, are unaware of how maternal morbidity can influence maternal health in the decades following pregnancy [108]. Because of this lack of awareness, patients often do not know to monitor their own health in the long term, and they may not receive referrals for proper follow-up services. The paucity of provider-focused technologies identified in this review may also present opportunities for the development of informatics-based interventions targeting physicians and affiliated providers—especially in the treatment and prevention of postpartum preeclampsia and hemorrhage—that could use technologies like decision support to improve maternal health.

XSL•FO
RenderX

Limitations

Our study has several limitations. Because this is a scoping review, we did not systematically assess the quality of individual studies. The studies reviewed may vary in quality or contain potential biases or methodological limitations. Therefore, this study is not intended to assess the quality of the evidence supporting informatics-based interventions on the given perinatal health conditions. Additionally, the nature of a scoping review is not to analytically aggregate studies so as to make claims about effect sizes. When additional, more comparable studies have accumulated, further research is needed to examine which informatics interventions are most effective and the size of the impact on maternal and perinatal health.

In addition, we only considered published studies that tested an informatics intervention and presented health-related outcomes. We excluded many studies because they tested feasibility or acceptability. It is possible that such excluded feasibility studies will eventually test health-related outcomes. Overall, this scoping review summarizes trends in the populations, geographic areas, technologies, and conditions targeted by informatics interventions tested and disseminated in the available literature.

Conclusions

This scoping review paints a picture of the landscape of informatics interventions aimed at preventing and treating maternal morbidity. Most interventions identified in this study were tested in North America, East Asia and the Pacific, or Western Europe. Most tested either smartphone/tablet applications or telehealth/telemedicine, and most technologies tested for new mental health conditions and GDM. Almost all the studies tested technologies on populations of women and reported medical, mental health, and patient behavior outcomes. Results suggest that there may be opportunities to use informatics technologies to target providers who treat women as well as conditions such as postpartum hemorrhage that are more likely to lead to mortality. As the landscape of informatics applications in health care continues to expand, maternal health is poised to be an important target of these interventions.

Acknowledgments

The authors would like to acknowledge the VA Medical Informatics Fellowship. This work was funded by the Department of Veterans Affairs Office of Academic Affiliations. The views expressed in this article are those of the authors and do not reflect the position or policy of the US Department of Veterans Affairs, the US government, or the authors' affiliated institutions.

Data Availability

The search strategies used in this study are openly available at searchRxiv [11,109-111].

Authors' Contributions

Conceptualization: JI (lead), DAH (equal), JCS (supporting)
Formal analysis: JI (lead), RCS (supporting), SS (supporting), DAH (supporting)
Supervision: DAH (lead), JI (equal)
Writing – original draft: JI (lead), JCS (supporting)
Writing – review and editing: JI (lead), JCS (supporting), DAH (supporting)
Methodology: JCS (lead), DAH (supporting)
Investigation: RCS (lead), SS (equal), SAEA (supporting), OO (supporting), BB (supporting)
Generative artificial intelligence was not used in the writing of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[DOCX File , 26 KB - ijmr_v14i1e64826_app1.docx]

References

1. Chivers BR, Boyle JA, Lang AY, Teede HJ, Moran LJ, Harrison CL. Preconception health and lifestyle behaviours of women planning a pregnancy: a cross-sectional study. J Clin Med 2020 Jun 02;9(6):1701 [FREE Full text] [doi: 10.3390/jcm9061701] [Medline: 32498329]
2. Bellamy L, Casas J, Hingorani AD, Williams DJ. Pre-eclampsia and risk of cardiovascular disease and cancer in later life: systematic review and meta-analysis. BMJ 2007 Nov 10;335(7627):974 [FREE Full text] [doi: 10.1136/bmj.39335.385301.BE] [Medline: 17975258]
3. Neiger R. Long-term effects of pregnancy complications on maternal health: a review. J Clin Med 2017 Jul 27;6(8):76. [doi: 10.3390/jcm6080076] [Medline: 28749442]

XSL•FO
RenderX

4. Ukah UV, Dayan N, Potter BJ, Ayoub A, Auger N. Severe maternal morbidity and risk of mortality beyond the postpartum period. Obstet Gynecol 2021 Feb 01;137(2):277-284. [doi: 10.1097/AOG.0000000000004223] [Medline: 33416296]
5. Robbins CL, Hutchings Y, Dietz PM, Kuklina EV, Callaghan WM. History of preterm birth and subsequent cardiovascular disease: a systematic review. Am J Obstet Gynecol 2014 Apr;210(4):285-297 [FREE Full text] [doi: 10.1016/j.ajog.2013.09.020] [Medline: 24055578]
6. Iftikhar P, Kuijpers MV, Khayyat A, Iftikhar A, DeGouvia De Sa M. Artificial intelligence: a new paradigm in obstetrics and gynecology research and clinical practice. Cureus 2020 Feb 28;12(2):e7124 [FREE Full text] [doi: 10.7759/cureus.7124] [Medline: 32257670]
7. Reed T. Health gaps emerge in how doctors and patients see AI. Axios. 2024 Apr 15. URL: https://www.axios.com/2024/04/15/doctors-patients-ai [accessed 2024-04-18]
8. Jean-Francois B, Bailey Lash T, Dagher RK, Green Parker MC, Han SB, Lewis Johnson T. The potential for health information technology tools to reduce racial disparities in maternal morbidity and mortality. J Womens Health (Larchmt) 2021 Feb;30(2):274-279 [FREE Full text] [doi: 10.1089/jwh.2020.8889] [Medline: 33211604]
9. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
10. Sampson M, McGowan J, Cogo E, Grimshaw J, Moher D, Lefebvre C. An evidence-based practice guideline for the peer review of electronic search strategies. J Clin Epidemiol 2009 Sep;62(9):944-952. [doi: 10.1016/j.jclinepi.2008.10.012] [Medline: 19230612]
11. Stumpff JC, Inderstrodt J, Haggstrom D. Systematic search: severe maternal morbidity and informatics (Engineering Village). searchRxiv Preprint posted online on May 11, 2023. [doi: 10.1079/searchrxiv.2023.00212]
12. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Implement 2021 Mar;19(1):3-10. [doi: 10.1097/XEB.0000000000000277] [Medline: 33570328]
13. Inderstrodt J. Maternal morbidity and informatics scoping review. Open Science Framework. 2024 Dec 20. URL: https://osf.io/nfvce [accessed 2025-02-13]
14. Identifying severe maternal morbidity (SMM). Centers for Disease Control and Prevention. 2024 May 15. URL: https://www.cdc.gov/maternal-infant-health/php/severe-maternal-morbidity/icd.html [accessed 2024-06-10]
15. Abbaspoor Z, Amani A, Afshari P, Jafarirad S. The effect of education through mobile phone short message service on promoting self-care in pre-diabetic pregnant women: a randomized controlled trial. J Telemed Telecare 2020 May;26(4):200-206. [doi: 10.1177/1357633X18791419] [Medline: 30193565]
16. Abejirinde IO, De Brouwere V, van Roosmalen J, van der Heiden M, Apentibadek N, Bardají A, et al. Viability of diagnostic decision support for antenatal care in rural settings: findings from the Bliss4Midwives Intervention in Northern Ghana. J Glob Health 2019 Jun;9(1):010420 [FREE Full text] [doi: 10.7189/jogh.09.010420] [Medline: 30937164]
17. Al-Ofi EA, Mosli HH, Ghamri KA, Ghazali SM. Management of postprandial hyperglycaemia and weight gain in women with gestational diabetes mellitus using a novel telemonitoring system. J Int Med Res 2019 Feb;47(2):754-764 [FREE Full text] [doi: 10.1177/0300060518809872] [Medline: 30442052]
18. Arias MP, Wang E, Leitner K, Sannah T, Keegan M, Delferro J, et al. The impact on postpartum care by telehealth: a retrospective cohort study. Am J Obstet Gynecol MFM 2022 May;4(3):100611 [FREE Full text] [doi: 10.1016/j.ajogmf.2022.100611] [Medline: 35331971]
19. Bartholomew ML, Soules K, Church K, Shaha S, Burlingame J, Graham G, et al. Managing diabetes in pregnancy using cell phone/internet technology. Clin Diabetes 2015 Oct;33(4):169-174 [FREE Full text] [doi: 10.2337/diaclin.33.4.169] [Medline: 26487790]
20. Baumel A, Tinkelman A, Mathur N, Kane JM. Digital peer-support platform (7Cups) as an adjunct treatment for women with postpartum depression: feasibility, acceptability, and preliminary efficacy study. JMIR Mhealth Uhealth 2018 Feb 13;6(2):e38 [FREE Full text] [doi: 10.2196/mhealth.9482] [Medline: 29439944]
21. Bellad MB, Goudar SS, Mallapur AA, Sharma S, Bone J, Charantimath US, CLIP India Working Group (Table S1). Community level interventions for pre-eclampsia (CLIP) in India: a cluster randomised controlled trial. Pregnancy Hypertens 2020 Jul;21:166-175 [FREE Full text] [doi: 10.1016/j.preghy.2020.05.008] [Medline: 32554291]
22. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in Norway. BMJ Open 2019 Nov 11;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
23. Carlisle N, Watson HA, Seed PT, Carter J, Kuhrt K, Tribe RM, et al. Impact of a medical mobile phone app (QUiPP) for predicting preterm birth on the anxiety and decisional conflicts faced by women in threatened preterm labour. Midwifery 2021 Jan;92:102864. [doi: 10.1016/j.midw.2020.102864] [Medline: 33137547]
24. Carroll AE, Biondich P, Anand V, Dugan TM, Downs SM. A randomized controlled trial of screening for maternal depression with a clinical decision support system. J Am Med Inform Assoc 2013;20(2):311-316 [FREE Full text] [doi: 10.1136/amiajnl-2011-000682] [Medline: 22744960]

XSL•FO
RenderX

25. Chan KL, Leung WC, Tiwari A, Or KL, Ip P. Using smartphone-based psychoeducation to reduce postnatal depression among first-time mothers: randomized controlled trial. JMIR Mhealth Uhealth 2019 May 14;7(5):e12794 [FREE Full text] [doi: 10.2196/12794] [Medline: 31094354]
26. Chappell L, Tucker K, Galal U, Yu LM, Campbell H, Rivero-Arias O, et al. Effect of self-monitoring of blood pressure on blood pressure control in pregnant individuals with chronic or gestational hypertension: the BUMP 2 randomized clinical trial. JAMA 2022 May 03;327(17):1666-1678 [FREE Full text] [doi: 10.1001/jama.2022.4726] [Medline: 35503345]
27. Cheung NW, Blumenthal C, Smith BJ, Hogan R, Thiagalingam A, Redfern J, et al. A pilot randomised controlled trial of a text messaging intervention with customisation using linked data from wireless wearable activity monitors to improve risk factors following gestational diabetes. Nutrients 2019 Mar 11;11(3):590 [FREE Full text] [doi: 10.3390/nu11030590] [Medline: 30862052]
28. Dennis C, Hodnett E, Kenton L, Weston J, Zupancic J, Stewart DE, et al. Effect of peer support on prevention of postnatal depression among high risk women: multisite randomised controlled trial. BMJ 2009 Jan 15;338:a3064 [FREE Full text] [doi: 10.1136/bmj.a3064] [Medline: 19147637]
29. Dennis C, Grigoriadis S, Zupancic J, Kiss A, Ravitz P. Telephone-based nurse-delivered interpersonal psychotherapy for postpartum depression: nationwide randomised controlled trial. Br J Psychiatry 2020 Apr;216(4):189-196. [doi: 10.1192/bjp.2019.275] [Medline: 32029010]
30. Felder JN, Segal Z, Beck A, Sherwood NE, Goodman SH, Boggs J, et al. An open trial of web-based mindfulness-based cognitive therapy for perinatal women at risk for depressive relapse. Cognitive Behav Pract 2017 Feb;24(1):26-37. [doi: 10.1016/j.cbpra.2016.02.002]
31. Ferrara A, Hedderson MM, Ching J, Kim C, Peng T, Crites YM. Referral to telephonic nurse management improves outcomes in women with gestational diabetes. Am J Obstet Gynecol 2012 Jun;206(6):491.e1-491.e5 [FREE Full text] [doi: 10.1016/j.ajog.2012.04.019] [Medline: 22631866]
32. Ferrara A, Hedderson MM, Brown SD, Albright CL, Ehrlich SF, Tsai A, et al. The comparative effectiveness of diabetes prevention strategies to reduce postpartum weight retention in women with gestational diabetes mellitus: the Gestational Diabetes' Effects on Moms (GEM) cluster randomized controlled trial. Diabetes Care 2016 Jan;39(1):65-74 [FREE Full text] [doi: 10.2337/dc15-1254] [Medline: 26657945]
33. Forsell E, Bendix M, Holländare F, Szymanska von Schultz B, Nasiell J, Blomdahl-Wetterholm M, et al. Internet delivered cognitive behavior therapy for antenatal depression: a randomised controlled trial. J Affect Disord 2017 Oct 15;221:56-64 [FREE Full text] [doi: 10.1016/j.jad.2017.06.013] [Medline: 28628768]
34. Garnweidner-Holme L, Henriksen L, Torheim LE, Lukasse M. Effect of the Pregnant+ smartphone app on the dietary behavior of women with gestational diabetes mellitus: secondary analysis of a randomized controlled trial. JMIR Mhealth Uhealth 2020 Nov 04;8(11):e18614 [FREE Full text] [doi: 10.2196/18614] [Medline: 33146620]
35. Ghaderi M, Farahani MA, Hajiha N, Ghaffari F, Haghani H. The role of smartphone-based education on the risk perception of type 2 diabetes in women with gestational diabetes. Health Technol 2019 Aug 7;9(5):829-837. [doi: 10.1007/s12553-019-00342-3]
36. Goetz M, Schiele C, Müller M, Matthies LM, Deutsch TM, Spano C, et al. Effects of a brief electronic mindfulness-based intervention on relieving prenatal depression and anxiety in hospitalized high-risk pregnant women: exploratory pilot study. J Med Internet Res 2020 Aug 11;22(8):e17593 [FREE Full text] [doi: 10.2196/17593] [Medline: 32780023]
37. Gong M, Zhang S, Xi C, Luo M, Wang T, Wang Y, et al. Comprehensive intervention during pregnancy based on short message service to prevent or alleviate depression in pregnant women: a quasi-experimental study. Early Interv Psychiatry 2021 Apr;15(2):352-359. [doi: 10.1111/eip.12953] [Medline: 32232971]
38. Guille C, Maldonado L, Simpson AN, Newman R, King C, Cortese B, et al. A non-randomized trial of in-person versus text/telephone screening, brief intervention and referral to treatment for pregnant and postpartum women. Psychiatr Res Clin Pract 2021;3(4):172-183 [FREE Full text] [doi: 10.1176/appi.prcp.20210027] [Medline: 36101654]
39. Guo H, Zhang Y, Li P, Zhou P, Chen L, Li S. Evaluating the effects of mobile health intervention on weight management, glycemic control and pregnancy outcomes in patients with gestational diabetes mellitus. J Endocrinol Invest 2019 Jun;42(6):709-714. [doi: 10.1007/s40618-018-0975-0] [Medline: 30406378]
40. Hantsoo L, Criniti S, Khan A, Moseley M, Kincler N, Faherty LJ, et al. A mobile application for monitoring and management of depressed mood in a vulnerable pregnant population. Psychiatr Serv 2018 Jan 01;69(1):104-107 [FREE Full text] [doi: 10.1176/appi.ps.201600582] [Medline: 29032705]
41. Hedderson MM, Brown SD, Ehrlich SF, Tsai A, Zhu Y, Quesenberry CP, et al. A tailored letter based on electronic health record data improves gestational weight gain among women with gestational diabetes mellitus: the Gestational Diabetes' Effects on Moms (GEM) cluster-randomized controlled trial. Diabetes Care 2018 Jul;41(7):1370-1377 [FREE Full text] [doi: 10.2337/dc17-1133] [Medline: 29669736]
42. Heller HM, Hoogendoorn AW, Honig A, Broekman BFP, van Straten A. The effectiveness of a guided internet-based tool for the treatment of depression and anxiety in pregnancy (MamaKits Online): randomized controlled trial. J Med Internet Res 2020 Mar 23;22(3):e15172 [FREE Full text] [doi: 10.2196/15172] [Medline: 32202505]

XSL•FO
RenderX

43. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007 Jun;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
44. Homko CJ, Deeb LC, Rohrbacher K, Mulla W, Mastrogiannis D, Gaughan J, et al. Impact of a telemedicine system with automated reminders on outcomes in women with gestational diabetes mellitus. Diabetes Technol Ther 2012 Jul;14(7):624-629 [FREE Full text] [doi: 10.1089/dia.2012.0010] [Medline: 22512287]
45. Hoppe KK, Thomas N, Zernick M, Zella JB, Havighurst T, Kim K, et al. Telehealth with remote blood pressure monitoring compared with standard care for postpartum hypertension. Am J Obstet Gynecol 2020 Oct;223(4):585-588. [doi: 10.1016/j.ajog.2020.05.027] [Medline: 32439388]
46. Huang F, Zhang S, Tian Y, Li L, Li Y, Chen X, et al. Effect of mobile health based peripartum management of gestational diabetes mellitus on postpartum diabetes: a randomized controlled trial. Diabetes Res Clin Pract 2021 May;175:108775 [FREE Full text] [doi: 10.1016/j.diabres.2021.108775] [Medline: 33771645]
47. Jannati N, Mazhari S, Ahmadian L, Mirzaee M. Effectiveness of an app-based cognitive behavioral therapy program for postpartum depression in primary care: a randomized controlled trial. Int J Med Inform 2020 Sep;141:104145. [doi: 10.1016/j.ijmedinf.2020.104145] [Medline: 32480319]
48. Klokkenga C, Enemark U, Adanu R, Lund S, Sørensen B, Attermann J. The effect of smartphone training of Ghanaian midwives by the Safe Delivery application on the incidence of postpartum hemorrhage: a cluster randomised controlled trial. Cogent Med 2019 Jul 17;6(1):1632016. [doi: 10.1080/2331205x.2019.1632016]
49. Krishnamurti T, Davis AL, Rodriguez S, Hayani L, Bernard M, Simhan HN. Use of a smartphone app to explore potential underuse of prophylactic aspirin for preeclampsia. JAMA Netw Open 2021 Oct 01;4(10):e2130804 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.30804] [Medline: 34714341]
50. Lanssens D, Vandenberk T, Smeets CJ, De Cannière H, Molenberghs G, Van Moerbeke A, et al. Remote monitoring of hypertension diseases in pregnancy: a pilot study. JMIR Mhealth Uhealth 2017 Mar 09;5(3):e25 [FREE Full text] [doi: 10.2196/mhealth.6552] [Medline: 28279948]
51. Lanssens D, Vonck S, Storms V, Thijs IM, Grieten L, Gyselaers W. The impact of a remote monitoring program on the prenatal follow-up of women with gestational hypertensive disorders. Eur J Obstet Gynecol Reprod Biol 2018 Apr;223:72-78 [FREE Full text] [doi: 10.1016/j.ejogrb.2018.02.015] [Medline: 29500948]
52. Latendresse G, Bailey E, Iacob E, Murphy H, Pentecost R, Thompson N, et al. A group videoconference intervention for reducing perinatal depressive symptoms: a telehealth pilot study. J Midwifery Womens Health 2021 Jan;66(1):70-77 [FREE Full text] [doi: 10.1111/jmwh.13209] [Medline: 33576146]
53. Lewey J, Murphy S, Zhang D, Putt ME, Elovitz MA, Riis V, et al. Effectiveness of a text-based gamification intervention to improve physical activity among postpartum individuals with hypertensive disorders of pregnancy: a randomized clinical trial. JAMA Cardiol 2022 Jun 01;7(6):591-599 [FREE Full text] [doi: 10.1001/jamacardio.2022.0553] [Medline: 35442393]
54. Li M, Lin Q, Shi J, Xi Y, Xiang C, Yong C, et al. The impact of lifestyle intervention on dietary quality among rural women with previous gestational diabetes mellitus-a randomized controlled study. Nutrients 2021 Jul 30;13(8):2642 [FREE Full text] [doi: 10.3390/nu13082642] [Medline: 34444802]
55. Lim K, Chan S, Lim SL, Tai BC, Tsai C, Wong SR, et al. A Smartphone App to Restore Optimal Weight (SPAROW) in women with recent gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 Mar 16;9(3):e22147 [FREE Full text] [doi: 10.2196/22147] [Medline: 33724204]
56. Mackillop L, Hirst JE, Bartlett KJ, Birks JS, Clifton L, Farmer AJ, et al. Comparing the efficacy of a mobile phone-based blood glucose management system with standard clinic care in women with gestational diabetes: randomized controlled trial. JMIR Mhealth Uhealth 2018 Mar 20;6(3):e71 [FREE Full text] [doi: 10.2196/mhealth.9512] [Medline: 29559428]
57. Milgrom J, Danaher BG, Gemmill AW, Holt C, Holt CJ, Seeley JR, et al. Internet cognitive behavioral therapy for women with postnatal depression: a randomized controlled trial of MumMoodBooster. J Med Internet Res 2016 Mar 07;18(3):e54 [FREE Full text] [doi: 10.2196/jmir.4993] [Medline: 26952645]
58. Milgrom J, Danaher BG, Seeley JR, Holt CJ, Holt C, Ericksen J, et al. Internet and face-to-face cognitive behavioral therapy for postnatal depression compared with treatment as usual: randomized controlled trial of MumMoodBooster. J Med Internet Res 2021 Dec 08;23(12):e17185 [FREE Full text] [doi: 10.2196/17185] [Medline: 34889742]
59. Miremberg H, Ben-Ari T, Betzer T, Raphaeli H, Gasnier R, Barda G, et al. The impact of a daily smartphone-based feedback system among women with gestational diabetes on compliance, glycemic control, satisfaction, and pregnancy outcome: a randomized controlled trial. Am J Obstet Gynecol 2018 Apr;218(4):453.e1-453.e7. [doi: 10.1016/j.ajog.2018.01.044] [Medline: 29425836]
60. 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 [FREE Full text] [doi: 10.1186/s12884-020-03341-9] [Medline: 33129314]
61. Ngai F, Wong PW, Leung K, Chau P, Chung K. The effect of telephone-based cognitive-behavioral therapy on postnatal depression: a randomized controlled trial. Psychother Psychosom 2015;84(5):294-303. [doi: 10.1159/000430449] [Medline: 26278623]

XSL•FO
RenderX

62. Nieminen K, Berg I, Frankenstein K, Viita L, Larsson K, Persson U, et al. Internet-provided cognitive behaviour therapy of posttraumatic stress symptoms following childbirth-a randomized controlled trial. Cogn Behav Ther 2016 Jun;45(4):287-306. [doi: 10.1080/16506073.2016.1169626] [Medline: 27152849]
63. Niksalehi S, Taghadosi M, Mazhariazad F, Tashk M. The effectiveness of mobile phone text massaging support for mothers with postpartum depression: a clinical before and after study. J Family Med Prim Care 2018;7(5):1058-1062 [FREE Full text] [doi: 10.4103/jfmpc.jfmpc_120_17] [Medline: 30598957]
64. Nishimwe A, Ibisomi L, Nyssen M, Conco DN. The effect of an mLearning application on nurses' and midwives' knowledge and skills for the management of postpartum hemorrhage and neonatal resuscitation: pre-post intervention study. Hum Resour Health 2021 Jan 26;19(1):14 [FREE Full text] [doi: 10.1186/s12960-021-00559-2] [Medline: 33499870]
65. Parsa S, Khajouei R, Baneshi MR, Aali BS. Improving the knowledge of pregnant women using a pre-eclampsia app: a controlled before and after study. Int J Med Inform 2019 May;125:86-90. [doi: 10.1016/j.ijmedinf.2019.03.001] [Medline: 30914185]
66. Perry H, Sheehan E, Thilaganathan B, Khalil A. Home blood-pressure monitoring in a hypertensive pregnant population. Ultrasound Obstet Gynecol 2018 Apr;51(4):524-530 [FREE Full text] [doi: 10.1002/uog.19023] [Medline: 29468771]
67. Posmontier B, Neugebauer R, Stuart S, Chittams J, Shaughnessy R. Telephone-administered interpersonal psychotherapy by nurse-midwives for postpartum depression. J Midwifery Womens Health 2016 Jul;61(4):456-466. [doi: 10.1111/jmwh.12411] [Medline: 26970401]
68. Potzel AL, Gar C, Banning F, Sacco V, Fritsche A, Fritsche L, et al. A novel smartphone app to change risk behaviors of women after gestational diabetes: a randomized controlled trial. PLoS One 2022;17(4):e0267258 [FREE Full text] [doi: 10.1371/journal.pone.0267258] [Medline: 35476681]
69. Rahman A, Akhter S, Nisha MK, Islam SS, Ashraf F, Rahman M, et al. Can mHealth improve access to safe blood for transfusion during obstetric emergency? Int J Womens Health 2017;9:235-243 [FREE Full text] [doi: 10.2147/IJWH.S120157] [Medline: 28461767]
70. Rasekaba TM, Furler J, Young D, Liew D, Gray K, Blackberry I, et al. Using technology to support care in gestational diabetes mellitus: quantitative outcomes of an exploratory randomised control trial of adjunct telemedicine for gestational diabetes mellitus (TeleGDM). Diabetes Res Clin Pract 2018 Aug;142:276-285. [doi: 10.1016/j.diabres.2018.05.049] [Medline: 29885390]
71. Sawyer A, Kaim A, Le H, McDonald D, Mittinty M, Lynch J, et al. The effectiveness of an app-based nurse-moderated program for new mothers with depression and parenting problems (eMums Plus): pragmatic randomized controlled trial. J Med Internet Res 2019 Jun 04;21(6):e13689 [FREE Full text] [doi: 10.2196/13689] [Medline: 31165715]
72. Shamshiri Milani H, Azargashb E, Beyraghi N, Defaie S, Asbaghi T. Effect of telephone-based support on postpartum depression: a randomized controlled trial. Int J Fertil Steril 2015;9(2):247-253 [FREE Full text] [doi: 10.22074/ijfs.2015.4246] [Medline: 26246884]
73. Silva-Jose C, Sánchez-Polán M, Diaz-Blanco Á, Coterón J, Barakat R, Refoyo I. Effectiveness of a virtual exercise program during COVID-19 confinement on blood pressure control in healthy pregnant women. Front Physiol 2021;12:645136 [FREE Full text] [doi: 10.3389/fphys.2021.645136] [Medline: 33776798]
74. Simhi M, Sarid O, Rowe H, Fisher J, Cwikel J. A cognitive-behavioral intervention for postpartum anxiety and depression: individual phone vs. group format. J Clin Med 2021 Dec 18;10(24):5952 [FREE Full text] [doi: 10.3390/jcm10245952] [Medline: 34945248]
75. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+ app) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]
76. St Pierre M, Luetcke B, Strembski D, Schmitt C, Breuer G. The effect of an electronic cognitive aid on the management of ST-elevation myocardial infarction during caesarean section: a prospective randomised simulation study. BMC Anesthesiol 2017 Mar 20;17(1):46 [FREE Full text] [doi: 10.1186/s12871-017-0340-4] [Medline: 28320312]
77. Sun Y, Li Y, Wang J, Chen Q, Bazzano AN, Cao F. Effectiveness of smartphone-based mindfulness training on maternal perinatal depression: randomized controlled trial. J Med Internet Res 2021 Jan 27;23(1):e23410 [FREE Full text] [doi: 10.2196/23410] [Medline: 33502326]
78. Sung J, Lee DY, Min KP, Park C. Peripartum management of gestational diabetes using a digital health care service: a pilot, randomized controlled study. Clin Ther 2019 Nov;41(11):2426-2434. [doi: 10.1016/j.clinthera.2019.09.005] [Medline: 31587813]
79. Sung Y, Zhang D, Eswaran H, Lowery CL. Evaluation of a telemedicine program managing high-risk pregnant women with pre-existing diabetes in Arkansas's Medicaid program. Semin Perinatol 2021 Aug;45(5):151421. [doi: 10.1016/j.semperi.2021.151421] [Medline: 34274150]
80. Tian Y, Zhang S, Huang F, Shi F, Li Y, Chen X, et al. Glycemic qualification rate and frequency of self-monitoring blood glucose glycemic qualification rate and frequency of self-monitoring blood glucose (SMBG) in women with gestational diabetes mellitus (GDM). Diabetes Res Clin Pract 2020 Dec;170:108482. [doi: 10.1016/j.diabres.2020.108482] [Medline: 32998018]

XSL•FO
RenderX

81. Tian Y, Zhang S, Huang F, Ma L. Comparing the efficacies of telemedicine and standard prenatal care on blood glucose control in women with gestational diabetes mellitus: randomized controlled trial. JMIR Mhealth Uhealth 2021 May 25;9(5):e22881 [FREE Full text] [doi: 10.2196/22881] [Medline: 33783365]
82. Triebwasser JE, Janssen MK, Hirshberg A, Srinivas SK. Successful implementation of text-based blood pressure monitoring for postpartum hypertension. Pregnancy Hypertens 2020 Oct;22:156-159. [doi: 10.1016/j.preghy.2020.09.001] [Medline: 32980623]
83. Tucker KL, Mort S, Yu L, Campbell H, Rivero-Arias O, Wilson HM, et al. Effect of self-monitoring of blood pressure on diagnosis of hypertension during higher-risk pregnancy: the BUMP 1 randomized clinical trial. JAMA 2022 May 03;327(17):1656-1665 [FREE Full text] [doi: 10.1001/jama.2022.4712] [Medline: 35503346]
84. Ugarriza DN, Schmidt L. Telecare for women with postpartum depression. J Psychosoc Nurs Ment Health Serv 2006 Jan;44(1):37-45. [doi: 10.3928/02793695-20060101-08] [Medline: 16475443]
85. Vandenberk T, Lanssens D, Storms V, Thijs IM, Bamelis L, Grieten L, et al. Relationship between adherence to remote monitoring and patient characteristics: observational study in women with pregnancy-induced hypertension. JMIR Mhealth Uhealth 2019 Aug 28;7(8):e12574 [FREE Full text] [doi: 10.2196/12574] [Medline: 31464190]
86. van den Heuvel JFM, Lely AT, Huisman JJ, Trappenburg JCA, Franx A, Bekker MN. SAFE@HOME: digital health platform facilitating a new care path for women at increased risk of preeclampsia - a case-control study. Pregnancy Hypertens 2020 Oct;22:30-36 [FREE Full text] [doi: 10.1016/j.preghy.2020.07.006] [Medline: 32717653]
87. van Montfort P, Smits LJM, van Dooren IMA, Lemmens SMP, Zelis M, Zwaan IM, et al. Implementing a preeclampsia prediction model in obstetrics: cutoff determination and health care professionals' adherence. Med Decis Making 2020 Jan;40(1):81-89 [FREE Full text] [doi: 10.1177/0272989X19889890] [Medline: 31789093]
88. Van Ryswyk EM, Middleton PF, Hague WM, Crowther CA. Postpartum SMS reminders to women who have experienced gestational diabetes to test for type 2 diabetes: the DIAMIND randomized trial. Diabet Med 2015 Oct;32(10):1368-1376. [doi: 10.1111/dme.12769] [Medline: 25816702]
89. Vigod SN, Slyfield Cook G, Macdonald K, Hussain-Shamsy N, Brown HK, de Oliveira C, et al. Mother matters: pilot randomized wait-list controlled trial of an online therapist-facilitated discussion board and support group for postpartum depression symptoms. Depress Anxiety 2021 Aug;38(8):816-825 [FREE Full text] [doi: 10.1002/da.23163] [Medline: 33949762]
90. Wozney L, Olthuis J, Lingley-Pottie P, McGrath PJ, Chaplin W, Elgar F, et al. Strongest Families™ Managing Our Mood (MOM): a randomized controlled trial of a distance intervention for women with postpartum depression. Arch Womens Ment Health 2017 Aug;20(4):525-537. [doi: 10.1007/s00737-017-0732-y] [Medline: 28593360]
91. Yang P, Lo W, He Z, Xiao X. Medical nutrition treatment of women with gestational diabetes mellitus by a telemedicine system based on smartphones. J Obstet Gynaecol Res 2018 Jul;44(7):1228-1234. [doi: 10.1111/jog.13669] [Medline: 29797375]
92. Yew TW, Chi C, Chan S, van Dam RM, Whitton C, Lim CS, et al. A randomized controlled trial to evaluate the effects of a smartphone application-based lifestyle coaching program on gestational weight gain, glycemic control, and maternal and neonatal outcomes in women with gestational diabetes mellitus: the SMART-GDM study. Diabetes Care 2021 Feb;44(2):456-463 [FREE Full text] [doi: 10.2337/dc20-1216] [Medline: 33184151]
93. Zera CA, Bates DW, Stuebe AM, Ecker JL, Seely EW. Diabetes screening reminder for women with prior gestational diabetes: a randomized controlled trial. Obstet Gynecol 2015 Jul;126(1):109-114 [FREE Full text] [doi: 10.1097/AOG.0000000000000883] [Medline: 26241263]
94. Zhuo Y, Pan Y, Lin K, Yin G, Wu Y, Xu J, et al. Effectiveness of clinical pharmacist-led smartphone application on medication adherence, insulin injection technique and glycemic control for women with gestational diabetes receiving multiple daily insulin injection: a randomized clinical trial. Prim Care Diabetes 2022 Apr;16(2):264-270. [doi: 10.1016/j.pcd.2022.02.003] [Medline: 35168915]
95. Najafi F, Hasani J, Izadi N, Hashemi-Nazari S, Namvar Z, Mohammadi S, et al. The effect of prepregnancy body mass index on the risk of gestational diabetes mellitus: a systematic review and dose-response meta-analysis. Obes Rev 2019 Mar;20(3):472-486. [doi: 10.1111/obr.12803] [Medline: 30536891]
96. Poorolajal J, Jenabi E. The association between body mass index and preeclampsia: a meta-analysis. J Matern Fetal Neonatal Med 2016 Nov;29(22):3670-3676. [doi: 10.3109/14767058.2016.1140738] [Medline: 26762770]
97. Blomberg M. Maternal obesity and risk of postpartum hemorrhage. Obstet Gynecol 2011 Sep;118(3):561-568. [doi: 10.1097/AOG.0b013e31822a6c59] [Medline: 21860284]
98. Cameron NA, Molsberry R, Pierce JB, Perak AM, Grobman WA, Allen NB, et al. Pre-pregnancy hypertension among women in rural and urban areas of the United States. J Am Coll Cardiol 2020 Dec 01;76(22):2611-2619 [FREE Full text] [doi: 10.1016/j.jacc.2020.09.601] [Medline: 33183896]
99. Osei E, Mashamba-Thompson TP. Mobile health applications for disease screening and treatment support in low-and middle-income countries: a narrative review. Heliyon 2021 Mar;7(3):e06639 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06639] [Medline: 33869857]

XSL•FO
RenderX

100. Peprah P, Abalo EM, Agyemang-Duah W, Budu HI, Appiah-Brempong E, Morgan AK, et al. Lessening barriers to healthcare in rural Ghana: providers and users' perspectives on the role of mHealth technology. A qualitative exploration. BMC Med Inform Decis Mak 2020 Feb 10;20(1):27 [FREE Full text] [doi: 10.1186/s12911-020-1040-4] [Medline: 32041608]
101. Ruhl LJ, Kiplagat J, O'Brien R, Wools-Kaloustian K, Scanlon M, Plater D, et al. A Global Health Reciprocal Innovation grant programme: 5-year review with lessons learnt. BMJ Glob Health 2023 Nov;8(Suppl 7):e013585 [FREE Full text] [doi: 10.1136/bmjgh-2023-013585] [Medline: 37977589]
102. Trost S, Beauregard J, Chandra G, Njie F, Berry J, Harvey A, et al. Pregnancy-related deaths: data from Maternal Mortality Review Committees in 36 US states, 2017-2019. CHAMPS. URL: https://champsonline.org/assets/files/Resources/ClinicalDocs/DiseaseCondition/MaternalMortalityReviewCommitteePregnancyRelatedDeathsData2017-2019.pdf [accessed 2024-01-15]
103. James AH, Federspiel JJ, Ahmadzia HK. Disparities in obstetric hemorrhage outcomes. Res Pract Thromb Haemost 2022 Jan;6(1):e12656 [FREE Full text] [doi: 10.1002/rth2.12656] [Medline: 35146237]
104. Four in 5 pregnancy-related deaths in the U.S. are preventable. Centers for Disease Control and Prevention. 2022 Sep 19. URL: https://www.cdc.gov/media/releases/2022/p0919-pregnancy-related-deaths.html [accessed 2024-06-10]
105. Ahn R, Gonzalez GP, Anderson B, Vladutiu CJ, Fowler ER, Manning L. Initiatives to reduce maternal mortality and severe maternal morbidity in the United States : a narrative review. Ann Intern Med 2020 Dec 01;173(11 Suppl):S3-S10 [FREE Full text] [doi: 10.7326/M19-3258] [Medline: 33253021]
106. Hailu EM, Maddali SR, Snowden JM, Carmichael SL, Mujahid MS. Structural racism and adverse maternal health outcomes: a systematic review. Health Place 2022 Nov;78:102923 [FREE Full text] [doi: 10.1016/j.healthplace.2022.102923] [Medline: 36401939]
107. DeNicola N, Grossman D, Marko K, Sonalkar S, Butler Tobah YS, Ganju N, et al. Telehealth interventions to improve obstetric and gynecologic health outcomes: a systematic review. Obstet Gynecol 2020 Feb;135(2):371-382 [FREE Full text] [doi: 10.1097/AOG.0000000000003646] [Medline: 31977782]
108. Roth H, LeMarquand G, Henry A, Homer C. Assessing knowledge gaps of women and healthcare providers concerning cardiovascular risk after hypertensive disorders of pregnancy-a scoping review. Front Cardiovasc Med 2019;6:178 [FREE Full text] [doi: 10.3389/fcvm.2019.00178] [Medline: 31850374]
109. Stumpff JC. Systematic search: severe maternal morbidity and informatics (Cochrane CENTRAL). searchRxiv Preprint posted online on September 27, 2022. [doi: 10.1079/searchRxiv.2022.00024]
110. Stumpff JC. searchRxiv Preprint posted online on September 6, 2022. [doi: 10.1079/searchRxiv.2022.00017]
111. Stumpff JC, Inderstrodt JM, Haggstrom D. Systematic search: severe maternal morbidity and informatics (IEEE Xplore). searchRxiv Preprint posted online on May 11, 2023. [doi: 10.1079/searchRxiv.2023.00211]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
GDM: gestational diabetes mellitus
HCAHPS: Hospital Consumer Assessment of Healthcare Providers and Systems
PCC: Population, Concept, and Context
PRESS: Peer Review of Electronic Search Strategies
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews
PTB: preterm birth
SMM: severe maternal morbidity

Edited by T de Azevedo Cardoso; submitted 27.07.24; peer-reviewed by A Morgan, R Appiah-Kubi; comments to author 01.10.24; revised version received 20.12.24; accepted 21.01.25; published 25.03.25.

Please cite as:
Inderstrodt J, Stumpff JC, Smollen RC, Sridhar S, El-Azab SA, Ojo O, Bowns B, Haggstrom DA
Informatics Interventions for Maternal Morbidity: Scoping Review
Interact J Med Res 2025;14:e64826
URL: https://www.i-jmr.org/2025/1/e64826
doi:10.2196/64826
PMID:

XSL•FO
RenderX

©Jill Inderstrodt, Julia C Stumpff, Rebecca C Smollen, Shreya Sridhar, Sarah A El-Azab, Opeyemi Ojo, Brendan Bowns, David A Haggstrom. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 25.03.2025. 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

Innovative, Technology-Driven, Digital Tools for Managing Pediatric Urinary Incontinence: Scoping Review

Lola Bladt[1,2], MSc; Jiri Vermeulen[1], MSc; Alexandra Vermandel[3,4], PhD; Gunter De Win[4,5], MD, PhD; Lukas Van Campenhout[2], PhD

[1]Department of Research and Development, Minze Health NV, Antwerp, Belgium

[2]Department of Product Development, Faculty of Design Sciences, University of Antwerp, Antwerp, Belgium

[3]Department of Rehabilitation Sciences and Physiotherapy, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

[4]Department of Urology, University Hospital Antwerp, Edegem, Belgium

[5]Antwerp Surgical Training, Anatomy and Research Centre (ASTARC), Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium

Corresponding Author:
Lola Bladt, MSc
Department of Product Development
Faculty of Design Sciences
University of Antwerp
Paardenmarkt 90/94
Antwerp, 2000
Belgium
Phone: 32 497848014
Email: lola.bladt@uantwerpen.be

Abstract

Background: Urinary incontinence affects approximately 7% to 10% of children during the day and 9% to 12% of children during the night. Treatment mainly involves lifestyle advice and behavioral methods, but motivation and adherence are low. Traditional tools such as pen-and-paper solutions may feel outdated and no longer meet the needs of today's "digital native" children. Meanwhile, digital interventions have already shown effectiveness in other pediatric health care areas.

Objective: This scoping review aimed to identify and map innovative, technology-driven, digital tools for managing pediatric urinary incontinence.

Methods: PubMed, Web of Science, and the Cochrane Library were searched in March 2022 without date restrictions, complemented by cross-referencing. Studies were eligible if they focused on pediatric patients (aged ≤18 years) with bladder and bowel dysfunctions and explored noninvasive, technology-based interventions such as digital health, remote monitoring, and gamification. Studies on adults, invasive treatments, and conventional methods without tangible tools were excluded. Gray literature was considered, but non–English-language, inaccessible, or result-lacking articles were excluded. A formal critical appraisal was not conducted as the focus was on mapping existing tools rather than evaluating effectiveness. Data analysis combined descriptive statistics and qualitative content analysis, categorizing tools through iterative coding and team discussions.

Results: In total, 66 articles were included, with nearly one-third (21/66, 32%) focusing on nocturnal enuresis. Our analysis led to the identification of six main categories of tools: (1) digital self-management (7/66, 11%); (2) serious games (7/66, 11%); (3) reminder technology (6/66, 9%); (4) educational media (12/66, 18%), further divided into video (5/12, 42%) and other media (7/12, 58%); (5) telehealth and remote patient monitoring (13/66, 20%), with subcategories of communication (5/13, 38%) and technological advances (8/13, 62%); and (6) enuresis alarm innovations (21/66, 32%), further divided into novel configurations (8/21, 38%) and prevoid alarms (13/21, 62%).

Conclusions: The field of pediatric urinary incontinence demonstrates a considerable level of innovation, as evidenced by the inclusion of 66 studies. Many tools identified in this review were described as promising and feasible alternatives to traditional methods. These tools were reported to enhance engagement, improve compliance, and increase patient satisfaction and preference while also having the potential to save time for health care providers. However, this review also identified gaps in research, highlighting the need for more rigorous research to better assess the tools' effectiveness and address the complex, multifaceted challenges of pediatric urinary incontinence management. Limitations of this review include restricting the search to 3 databases,

XSL•FO
RenderX

excluding non–English-language articles, the broad scope, and single-reviewer screening, although frequent team discussions ensured rigor. We propose that future tools should integrate connected, adaptive, and personalized approaches that align with stakeholder needs, guided by a multidisciplinary, human-centered framework combining both qualitative and quantitative insights.

(Interact J Med Res 2025;14:e66336) doi:10.2196/66336

KEYWORDS

pediatric urinary incontinence; nocturnal enuresis; behavioral therapy; urotherapy; patient compliance; digital health; serious games; telehealth; health technology; enuresis alarm; artificial intelligence; AI

Introduction

Background

Urinary incontinence (UI) is prevalent among children, affecting approximately 7% to 10% (aged 5-13 years) during the day [1] and 9% to 12% (aged 5-15 years) during the night [2-4]. UI often leads to low self-esteem, shame, and stress, negatively impacting the child's psychosocial well-being and overall quality of life [5-9]. The primary treatment approach for pediatric UI, once anatomical and neurogenic causes are excluded, mainly involves lifestyle advice and behavioral methods [10-12]. These methods vary depending on whether the UI occurs during the day or at night.

For daytime UI (DUI), the recommended first-line approach is standard urotherapy. It encompasses patient education, keeping track of symptoms through bladder diaries, regulating fluid intake and toilet routines, adhering to scheduled voiding, and providing support and encouragement [13-16]. An important part of this treatment is teaching children appropriate responses to their bladder signals. Standard urotherapy achieves success rates ranging from 40% to 56% [17-19]. For nighttime UI, also known as nocturnal enuresis (NE), one of the recommended first-line treatments is NE alarm training [20]. This method uses a moisture sensor that triggers an alarm to wake the child upon bed-wetting. Through this approach, children are conditioned to recognize and respond to bladder signals, ultimately helping reduce bed-wetting episodes. The success rate of alarm training is between 50% and 70% [21]. However, motivation and adherence, both crucial for these behavioral methods to be successful [22,23], are low in practice [22,24,25].

There are many factors that influence motivation and adherence [15,24-27], including duration of therapy, required effort, symptom bother, and awareness but also the supporting tools' usability and acceptance. The evaluation and management of pediatric UI traditionally rely on pen-and-paper bladder diaries [28-30], along with verbal and written instructions for behavioral techniques [16] and traditional NE alarms [31]. However, these analogue methods may seem outdated to today's "digital native" children, who are accustomed to using technology and digital tools [32]. Consequently, traditional approaches to pediatric UI management may no longer meet today's children's needs and expectations, potentially affecting motivation and adherence. Meanwhile, digital interventions have already proven effective as alternatives to traditional methods in other pediatric health care areas such as pain tracking [33], attention-deficit/hyperactivity disorder therapy [34], childhood obesity prevention [35], and asthma management [36].

Objectives

Therefore, our objective was to identify and map innovative, technology-driven, digital tools for pediatric UI management as alternatives to conventional approaches such as pen-and-paper solutions. In addition, we aimed to provide an overview of these tools' characteristics, including their features, design drivers, and any reported evidence supporting their use.

Methods

This scoping review followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines as outlined by Tricco et al [37] (see Multimedia Appendix 1 for the PRISMA-ScR checklist).

Eligibility Criteria

Eligibility criteria were determined based on the population, concept, and context framework. The *population* for this review included pediatric patients experiencing conditions such as bladder and bowel dysfunction (BBD), NE, overactive bladder (OAB), underactive bladder, and dysfunctional voiding (DV), as well as neurogenic conditions such as spina bifida. Studies involving adult populations (aged ≥18 years) or unrelated conditions (eg, cancer or rare diseases) were excluded.

The *concept* focused on innovative, technology-driven, digital tools for managing pediatric UI and related dysfunctions. Eligible tools included smartphone apps, multimedia resources, and remote monitoring technologies, whereas conventional methods or treatment plans without tangible tools were excluded.

The *context* was limited to studies that addressed noninvasive therapeutic approaches, particularly behavioral methods such as urotherapy, bladder training, and alarm training. Invasive approaches, including surgeries and their associated tools, were excluded.

All types of literature were considered, including gray literature (eg, conference proceedings, preprints, and technical reports), to provide a broad range of available literature and reduce publication bias. However, articles not presenting results or findings, not written in English, or lacking full-text availability were excluded.

Search Strategy

The search (see Multimedia Appendix 2 for the full search strategy) was conducted in March 2022 across 3 databases—PubMed, Web of Science, and the Cochrane Library—with no limitations on publication dates. The

XSL•FO
RenderX

population, concept, and context framework guided the selection of search terms.

Key search terms for *population* included "children," "adolescents," "pediatric," "bladder bowel dysfunction," and "enuresis." For *concept*, terms such as "digital," "mHealth," "gamification," "multimedia," and "remote monitoring" were used. *Context*-related search terms included "urotherapy," "bladder training," "alarm therapy," "self-monitoring," and "behavior management."

Study Selection

The study selection process was conducted by the first author (LB), with consultation and supervision provided by the team of authors. After removing duplicates, potential studies underwent screening based on their titles and abstracts in line with the eligibility criteria, with any uncertainties discussed with the author team.

Subsequently, LB read and assessed the full-text articles, periodically consulting with the team to ensure consistency and rigor. To reduce selection bias from database choice, we conducted cross-referencing, scanning the reference lists of the included articles to identify additional relevant studies.

The selected studies underwent final review and confirmation by the entire author team. A formal critical appraisal or methodological quality assessment was not conducted as the aim of this scoping review was to identify and map the available evidence on innovative, technology-driven, digital tools rather than draw formal conclusions on their effectiveness.

Data Extraction

Data were extracted and charted using Microsoft Excel (Microsoft Corp) according to four main heading groups: (1) article information, (2) participant characteristics (population), (3) tool description (concept), and (4) research methods and results (context). Article information encompassed details such as article title, authors, journal, publication year, country, and database source. Participant characteristics included sample size, age range, sex distribution, and dysfunction indication. Tool description covered intended therapy type, tool design details, and its design drivers. Finally, research methods and results comprised study design, intervention groups, study outcomes, and conclusions.

Data Analysis

A combination of descriptive statistics and basic qualitative content analysis was used for data analysis. Descriptive statistics were used to analyze study characteristics, providing frequency counts and proportions for extraction items such as study design and dysfunction indication. Basic qualitative content analysis was used to identify and classify innovative, technology-driven, digital tools into distinct categories. The process began with multiple readings of the extraction items related to the tool description to familiarize ourselves with the data. Following an inductive approach, initial notes were made to generate codes, and coding was manually conducted in Microsoft Excel, where relevant terms were highlighted and grouped into emerging patterns. Through iterative refinement and team discussions, codes were merged, removed, or redefined to eliminate overlap and ensure sufficient distinction. The final set of codes was organized into 6 broad categories, 3 of which were further subdivided to capture more specific aspects. LB conducted the preliminary analysis, and to ensure consistency and reliability, the coding process was reviewed and refined collaboratively by the coauthors at multiple stages. Any disagreements on categorization were addressed through discussions to achieve consensus.

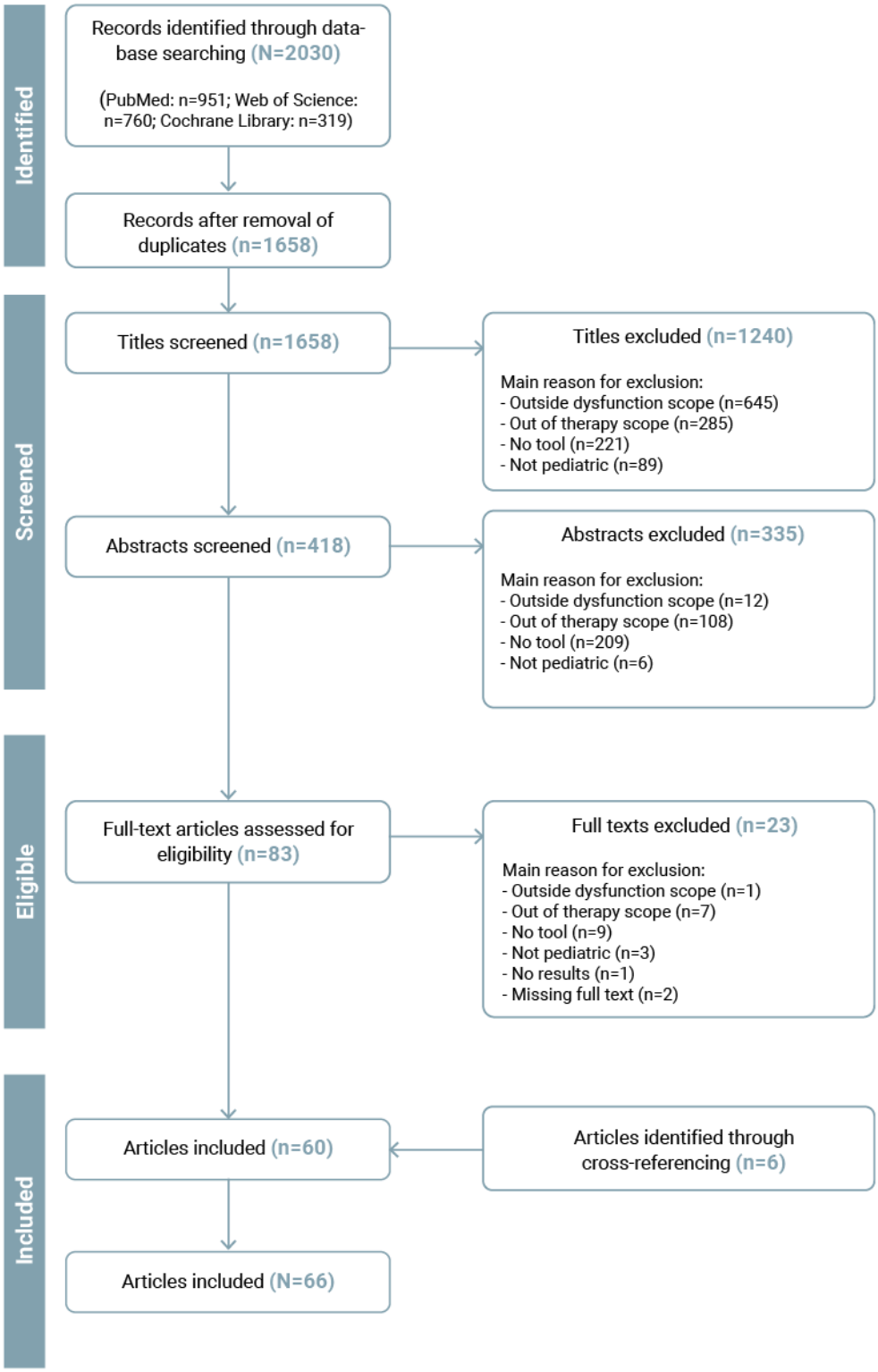
Results

Retrieval of Studies

The search retrieved a total of 2030 articles (PubMed: n=951, 46.85%; Web of Science: n=760, 37.44%; Cochrane Library: n=319, 15.71%). After removing duplicates (372/2030, 18.33%), a total of 81.67% (1658/2030) of the studies were included in the title and abstract screening. Following title and abstract screening, of the 1658 articles, 1575 (94.99%) were excluded, leaving 83 (5.01%) for full-text evaluation. Finally, 60 articles were included in the data extraction process, with an additional 6 identified through cross-referencing, resulting in a total of 66 included articles. Figure 1 illustrates the flowchart of the screening and selection process.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the literature screening and selection process.



General Analysis to Describe the Study Characteristics

Of the included studies (n=66), most were randomized controlled trials (RCTs; 14/66, 21%) and validation studies (13/66, 20%), followed by quasi-experiments (10/66, 15%), prospective studies with no control group (7/66, 11%), and qualitative or mixed methods studies (8/66, 12%), as shown in Table 1. Nearly one-third (21/66, 32%) of the identified tools targeted children with NE, followed by DV (9/66, 14%), BBD (7/66, 11%), and DUI (7/66, 11%). More than one-third (25/66, 38%) of the studies evaluated outcomes related to patient experience, such as satisfaction, motivation, and usability. A mean sample size of 82 (SD 83; range 14-353) participants was found for the RCTs, quasi-experiments, and prospective studies (no control), indicating a large variation in sample size.

We identified and established a total of six primary categories of innovative, technology-driven, digital tools (Figure 2): (1) digital self-management (7/66, 11%), (2) serious games (7/66,

XSL•FO
RenderX

11%), (3) reminder technology (6/66, 9%), (4) educational media (12/66, 18%), (5) telehealth and remote patient monitoring (RPM; 13/66, 20%), and (6) enuresis alarm innovations (21/66, 32%). Further subcategories were identified within educational media, including video (5/12, 42%) and other media (7/12, 58%); within telehealth and RPM, including communication (5/13, 38%) and technological advances (8/13, 62%); and within enuresis alarm innovations, including novel configurations (8/21, 38%) and prevoid alarms (13/21, 62%).

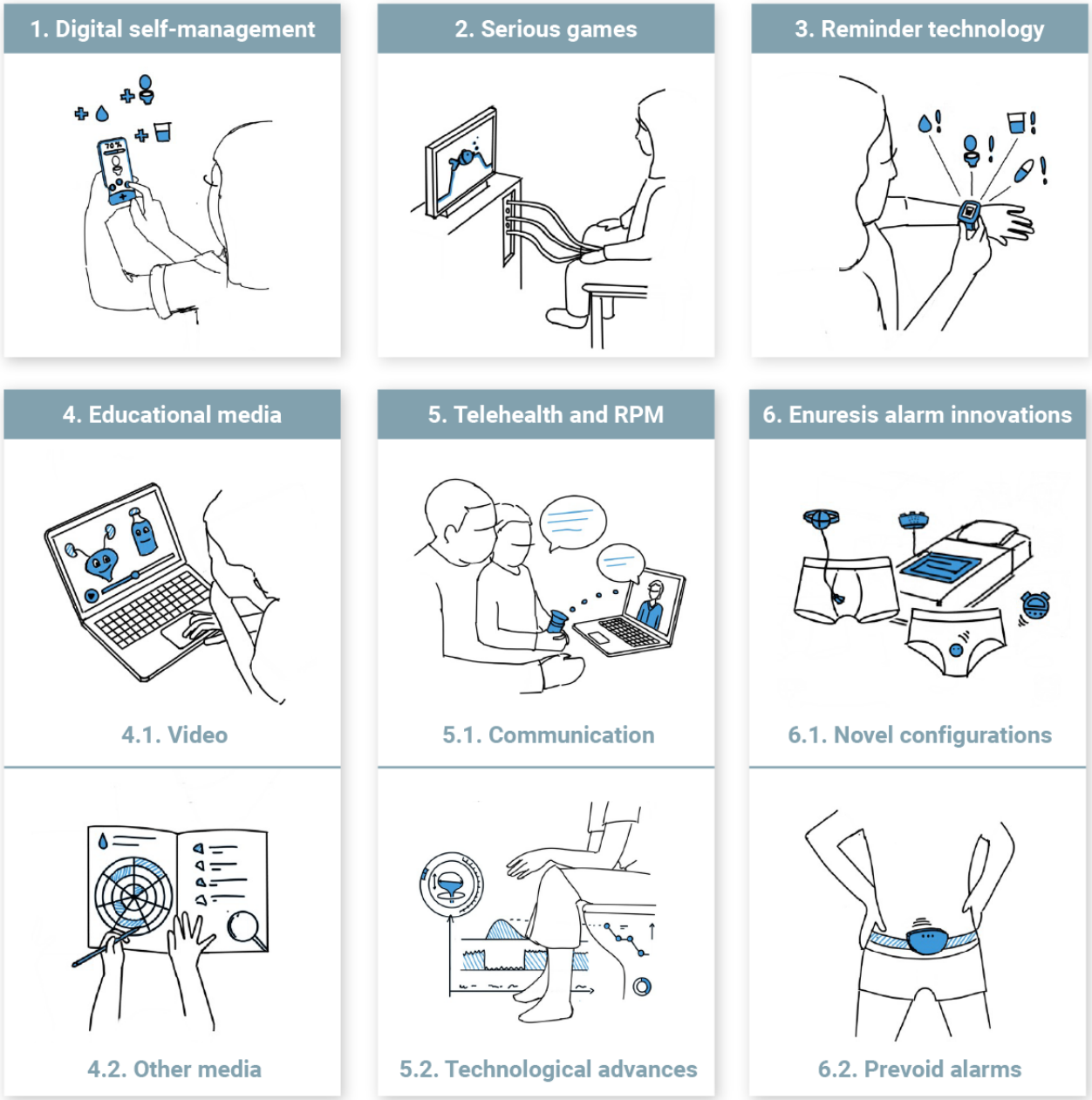
Table 1. Summary of study characteristics (n=66).

Study characteristic	Studies, n (%)
Study design	
RCT[a]	14 (21)
Feasibility or validation study	13 (20)
Quasi-experiment	10 (15)
Prospective nc[b]	7 (11)
Review	6 (9)
Mixed methods study	5 (8)
Retrospective study	4 (6)
Qualitative study	3 (5)
Modeling	2 (3)
Case study	1 (2)
Design-based research	1 (2)
Dysfunction indication	
NE[c]	21 (32)
DV[d]	9 (14)
BBD[e]	7 (11)
DUI[f]	7 (11)
DUI or nighttime UI[g]	5 (8)
Neurogenic	5 (8)
UTI[h]	4 (6)
Developmental or neurodevelopmental disorder	3 (5)
OAB[i]	2 (3)
UAB[j]	1 (2)
Bowel dysfunction	1 (2)
Anatomical	1 (2)

[a]RCT: randomized controlled trial.
[b]nc: no control.
[c]NE: nocturnal enuresis.
[d]DV: dysfunctional voiding.
[e]BBD: bladder and bowel dysfunction.
[f]DUI: daytime urinary incontinence.
[g]UI: urinary incontinence.
[h]UTI: urinary tract infection.
[i]OAB: overactive bladder.
[j]UAB: underactive bladder.

XSL•FO
RenderX

Figure 2. Overview of the 6 categories and subcategories of innovative, technology-driven, digital tools. RPM: remote patient monitoring.



Overall, we found that the distribution of dysfunction types was generally well balanced across the different tool categories. Digital self-management exhibited a diverse distribution, addressing a range of dysfunctions including BBD (1/7, 14%), DUI or nighttime UI (1/7, 14%), and neurogenic disorders (2/7, 29%). Serious games were predominantly used for DV (5/7, 71%). Reminder technology was used for timed voiding in cases of DUI and OAB (3/6, 50%), as well as for self-management of neurogenic and developmental or neurodevelopmental disorders (3/6, 50%). Educational media was used across a range of dysfunctions, with slightly more use for NE (4/12, 33%) and urinary tract infections (UTIs; 3/12, 25%). Telehealth was primarily used for BBD (4/13, 31%) and DUI or nighttime UI (4/13, 31%), with additional use also observed for DV (3/13, 23%). Finally, enuresis alarm innovations were predominantly associated with NE (16/21, 76%), as expected, although not exclusively. They were also applied for other conditions, such as DUI (2/21, 10%) and developmental or neurodevelopmental disorders (2/21, 10%).

Categories of Innovative, Technology-Driven, Digital Tools

Digital Self-Management

Category Overview and Key Drivers

Digital devices such as smartphones, tablets, and computers have become increasingly prevalent in our society, especially among children and adolescents [38]. Therefore, they offer a practical and accessible medium with a range of health care possibilities, such as digital self-management to enhance patient engagement and improving access, availability, and continuity of care [38,39].

XSL•FO
RenderX

Tools Within This Category

This review identified 11% (7/66) of the studies on digital self-management, with extended details available in Table S1 in Multimedia Appendix 3 [40-105]. These included electronic bladder diary apps [40,41], which served as digital alternatives to traditional paper versions to improve patient compliance. In addition, 3 self-management applications with training programs were included, targeting conditions such as DUI (smartphone app) [42], bowel dysfunctions (internet application) [43], and spina bifida (smartphone app) [44,45]. These self-management applications incorporated features such as weekly modules with exercises; daily goals; self-monitoring of parameters such as fluid intake, toilet visits, medication, and mood status; reminders; progress tracking; personalized feedback; and rewards such as in-application trophies and stars. Furthermore, a parent software portal [46] was identified, enabling parents to monitor their children's symptoms, such as UTIs, to aid clinical decision-making.

Study Results and Conclusions

The studies on digital self-management tools reported feasibility, acceptability, ease of use, and patient and parent satisfaction [40-46]. Furthermore, they described improvements in patient compliance, as observed in both the bowel dysfunction self-management program and the parent software portal, with high completion rates and low dropout [43,46]. However, the parent software portal was described as facing challenges in effectively delivering information to health care professionals, lacking integration into their busy clinical practice [46].

Despite many promising aspects, the studies often reported limitations in providing robust evidence of effectiveness. For example, the electronic bladder diary format was described as offering data quality advantages for certain patients but was not found to be superior to a paper-based bladder diary [40]. The sole RCT investigating a self-management program for bowel dysfunctions reported greater improvements in patients' quality of life and reduction in overall symptom severity. However, this was in comparison to a nonequivalent control group of children on a waitlist who were not receiving any treatment [43].

Serious Games

Category Overview and Key Drivers

Serious games are interactive applications designed as games but intended for educational or therapeutic purposes. By blending elements of fun and seriousness, serious games offer a playful way for children to learn about complex topics such as bladder and bowel health. This approach can be particularly effective for engaging children, who have limited attention spans and may struggle with processing information. Moreover, the entertainment aspect of serious games is crucial for motivating and engaging children as they strive for success within the game [41,106,107].

Tools Within This Category

Table S2 in Multimedia Appendix 3 [40-105] shows that serious games were mainly used for pelvic floor biofeedback training through biogames (6/7, 86%) [47-52]. In these biogames, the game action was controlled by the appropriate muscular contractions and relaxations detected by pelvic floor electromyography (EMG) patches worn by the child. Various animated scenarios were available, such as golfing, flying spaceships, riding an elephant, playing basketball, and swimming with dolphins or fish. In addition, one serious game was found for standard urotherapy, which was played on a smartphone [53]. In this game, the child was rewarded for maintaining a daily bladder diary and voiding adequately during the 3-month training program, with additional rewards for staying dry.

Study Results and Conclusions

Overall, the studies described positive effects of biogames for pelvic floor biofeedback training on health outcomes, particularly for children with DV [47-50,52] and, in one study, for those with underactive bladder [51]. However, understanding the precise impact of adding a game to the therapy is challenging as the studies reported limitations such as retrospective designs [49], lack of control groups [47], and nonequivalent therapy comparisons [48,50,51].

Findings varied across the studies. One study reported similar results in half the time with biogames (mean of 3.6, SD 1.8 sessions) compared to without (mean of 7.6, SD 5.2 sessions) [49], whereas another study reported a relatively high number of treatment sessions with biogames (mean of 9.6, SD 1.3 sessions) [50]. In the sole RCT comparing equivalent therapies [52], no differences were observed in terms of clinical success, cooperation, and motivation between pelvic floor biofeedback training with and without biogames. Nevertheless, subjective observations suggested that using biogames effectively captured children's attention and prevented boredom during the sessions, particularly among younger children [52].

Similarly, in the comparison of standard urotherapy with and without a serious game, no significant differences were reported in training results, quality of life, or motivation [53]. However, children expressed a clear preference for the game. They provided valuable suggestions for its improvement, such as introducing more dynamic gameplay with changing storylines and evolving game types, along with features such as regular updates on training progress and detailed information about their bladder problem.

Reminder Technology

Category Overview and Key Drivers

Reminder technology uses programmable timers to activate alarms aimed at assisting children in adhering to therapy or self-management schedules without requiring constant prompting from parents. This approach supports children's independence as parental reminders may sometimes contribute to resistance, with children preferring autonomy in adhering to their schedules [108].

Tools Within This Category

A total of 9% (6/66) of the studies were categorized under reminder technology (Table S3 in Multimedia Appendix 3 [40-105]). Half (3/6, 50%) of these studies evaluated timer watches, which were used in standard urotherapy to assist with

XSL•FO
RenderX

timed voiding [54-56]. The timer watch can be programmed to automatically alarm at specified intervals, such as every 2 hours, helping children maintain a consistent voiding schedule.

The other half (3/6, 50%) of the studies focused on self-management reminders for children with neurogenic or developmental or neurodevelopmental disorders [57-59]. A laptop calendar program reminded children with prospective memory disabilities to complete and log daily schedules, such as self-catheterization and medication, through visual and auditory modalities [59]. A smartwatch with a built-in fitness tracker displayed visual icons to prompt children with spina bifida to engage in various daily activities from self-care to playtime and rewarding timely completion with stars [58]. Finally, "Potty Monkey," an interactive toy, prompted children with special needs to void at set times, providing reminders, modeling techniques, and positive reinforcement for time-based toilet training [57].

Study Results and Conclusions

Reminder technology for timed voiding using a timer watch was reported to be superior in treating severe DUI in children with OAB compared to standard urotherapy without it, with improved adherence to the voiding schedule playing a key factor [55,56]. In addition, another study described reminder technology as an equally effective but simpler alternative to enuresis alarms (using moisture sensors) in treating DUI, also reporting good adherence rates [54].

For self-management in children with neurogenic or developmental or neurodevelopmental disorders, reminder technology was reported to be feasible, although the findings were based on small sample sizes [57-59]. The tools were described as alleviating anxiety in caregivers and patients by providing a log of tasks [59] and fostering child independence through clear routines [58]. However, patient acceptance varied, with the smartwatch being highly accepted [58] and the interactive toy receiving lower acceptance [57], with aesthetics and flexibility playing a key role.

Educational Media

Video

Category Overview and Key Drivers

Videos, as educational media, offer a visually engaging and interactive format for explaining complex concepts and techniques to patients and parents. They provide a convenient and 24/7-available alternative or supplement to in-person visits. Moreover, videos offer anonymity, which can be beneficial for individuals who feel embarrassed or hesitant to seek medical advice in person. As a result, the internet and video-sharing platforms have become popular sources of health-related information [32].

Tools Within This Category

In our review, 8% (5/66) of the studies were classified under the subcategory of video within educational media (Table S4 in Multimedia Appendix 3 [40-105]). These videos covered a range of topics. Some supported behavioral therapy, such as an animated bladder training video [60,61] and an abundance of YouTube videos on NE therapy [62,63]. Another video focused on preparing children for surgical interventions, featuring actual hospital staff and using role-play modeling techniques to demonstrate the anesthesia process [64].

Study Results and Conclusions

Multiple studies within this subcategory (3/5, 60%) described videos as effective educational media for patients offering alternatives to in-person strategies [60,61,64]. In a noninferiority RCT, the animated bladder training video was found to be equivalent to standard urotherapy in reducing bladder and bowel symptoms [60] and improving quality of life scores [61]. Similarly, the presurgery video was reported to effectively reduce preoperative anxiety and resulted in less postoperative maladaptive behavior [64].

While videos were described as effective tools for educating patients and families, caution is recommended when relying on video-sharing platforms such as YouTube that lack robust systems to ensure the accuracy, completeness, and quality of the information [62,63]. The studies indicated a need for a balance between adherence to evidence-based guidelines and the creation of engaging content that captures and maintains viewer attention.

Other Media

Category Overview and Key Drivers

Effective information delivery is crucial for raising awareness, setting expectations, promoting patient adherence, and ensuring data quality. Unfortunately, verbal explanations by physicians may be poorly understood due to time constraints and complexity, whereas written information may not always be suitable or desirable for children, resulting in limited child participation. Thus, using other types of educational media tailored and engaging for children may help ensure effective information delivery to pediatric patients.

Tools Within This Category

A total of 11% (7/66) of the studies were classified under other types of educational media (Table S5 in Multimedia Appendix 3 [40-105]). This category included 2 homework books for UI: one that used coloring activities to log the severity of incontinence [65] and another that featured practical exercises for standard urotherapy [66]. In addition, a multimedia program for NE was identified, which combined text, drawings, cartoons, and sound to educate children on bed-wetting [67,68]. Finally, educational materials on UTIs included educational urine kits for clean-catch urine collection [69-71] and a low-cost origami-based microscope that used a mobile phone camera to raise awareness of genital hygiene and UTIs in schools [71].

Study Results and Conclusions

The UI homework books were reported to be feasible, usable, and well received, with good compliance [65,66]. In addition, the homework book for standard urotherapy was associated with increased fluid intake and hold time [66].

The multimedia program for NE was noted to show promise in holding children's attention and improving their knowledge, especially for younger children with low literacy levels [67]. However, results from the RCT indicated that, while the multimedia program was not more effective than traditional

XSL•FO
RenderX

verbal and written information, it could help save health care professionals time in patient education [68].

Similarly, urine kits with educational pamphlets for clean-catch urine collection did not result in a significant reduction in contamination rates but may help streamline patient education [69,70]. In addition, the origami-based microscope was described as a promising educational tool for promoting health awareness by detecting UTI-positive urine samples reasonably well [71].

Telehealth and RPM

Communication

Category Overview and Key Drivers

Telehealth and RPM enable remote communication and data exchange between patients and clinicians. Although face-to-face communication has been shown to improve treatment adherence [27,109,110], constraints such as time, cost, and travel can lead to extended intervals between consultations, potentially affecting patient motivation [111]. Telehealth may address some of these issues by offering consistent follow-up visits and motivational support without the need for physical visits.

Tools Within This Category

A total of 8% (5/66) of the studies were classified within the subcategory of communication using telehealth and RPM (Table S6 in Multimedia Appendix 3 [40-105]). One study explored videoconferencing for remote patient visits [72], whereas another study evaluated a UTI home monitoring system [73]. This system facilitated communication between parents of children with neurogenic bladder and urology nursing staff to alert them of potential UTI signs. It comprised a portable urinalysis device, Bluetooth weight scale, thermometer, electronic bladder diary, and mobile communication device. In addition, an embodied conversational agent, "Dr. Evie," was identified, offering personalized lifestyle advice and empathetic coaching for UI self-management while generating reports for general practitioners [74-76].

Study Results and Conclusions

The studies reported high patient and parent satisfaction with telehealth, citing reduced stress and increased convenience, especially for rural families, due to decreased travel time [72-76]. In addition, the UI self-management program featuring an embodied conversational agent was reported to improve patient adherence rates from 50% to 76% on average and alleviate symptoms [74-76]. Advice from nonfamily members and a sense of alliance with Dr Evie were noted to enhance program engagement.

Health care providers expressed mixed feelings about telehealth, acknowledging improved access to care but reporting technical barriers in conducting physical examinations and establishing rapport, particularly with pediatric patients [72]. Integrating RPM systems such as the UTI monitoring system may mitigate these technical barriers by providing valuable data for remote support and enhancing patient and parent self-efficacy [73]. However, patients' selective use of certain products, mainly the urinalysis device, was noted to create data gaps that hinder clinical evaluation. In addition, technical issues contributed to increased workload for clinical teams, emphasizing the importance of careful consideration of all stakeholders' diverse needs and preferences to facilitate successful integration into clinical practice.

Technological Advances

Category Overview and Key Drivers

Technological advances enabling telehealth and RPM aim to overcome limitations of in-clinic procedures, including the psychological stress that children often experience in hospital settings [77,112]. These procedures can also be inconvenient and time-consuming, as seen with noninvasive uroflowmetry, which requires voiding with a full bladder [78]. Similarly, therapies such as EMG biofeedback demand significant time and financial resources due to poor compliance and the need for multiple visits [77]. Meanwhile, invasive studies carry health risks and are often discouraged, especially for young patients [113,114].

Tools Within This Category

In total, 12% (8/66) of the studies were classified in the telehealth and RPM subcategory that focused on technological advances (Table S7 in Multimedia Appendix 3 [40-105]). These ranged from a machine learning algorithm to predict enuresis in children [79] to a noninvasive alternative to traditional invasive urodynamics using near-infrared spectrometry [80].

Advances in uroflowmetry beyond traditional gravimetric methods for home use were also reviewed [78]. These included basic volumetric, mechanical, and electrical methods such as capacitance sensors. Unconventional approaches were also discussed, including image processing for drop patterns, ultrasound Doppler effect, vibration sensors on the penis, and acoustic technology using a smartphone app to record urine sound [78,81].

Several studies (5/8, 63%) explored adapting existing technologies for home use. This included gravimetric uroflowmetry for home use [77,81] and its application in therapy settings to provide biofeedback during voiding [77,82]. Another study introduced home pelvic floor biofeedback through a wearable EMG device featuring a watchlike design with a light scale to indicate muscle activity strength [83]. In addition, transcutaneous electrical nerve stimulation was evaluated as a user-friendly substitute for in-clinic percutaneous tibial nerve stimulation to enhance continence, enabling self-administration at home without invasive needle insertion [84].

Study Results and Conclusions

Overall, the studies described promising advances in technology. The machine learning technology was described as having promising accuracy in screening for enuresis, with older age at toilet training initiation identified as a significant predictor [79]. Both uroflowmetry and EMG biofeedback were described as effective in reducing symptoms in patients with DV [82,83], with home uroflowmetry biofeedback showing promise and patient preference [77]. Transcutaneous electrical nerve stimulation was reported to show efficacy in reducing NE episodes, providing a safe and noninvasive at-home treatment option [84].

XSL•FO
RenderX

However, some novel technologies such as near-infrared spectrometry were described as promising and well accepted by patients but were also noted to lack standardized benchmarks and face challenges in ensuring consistent and reliable data that align with established norms and guidelines [80]. Similarly, while various uroflowmetry methods exist, gravimetric uroflowmeters were favored due to their simplicity and reliability [78]. For example, sound-based uroflowmetry, despite showing strong correlations with conventional uroflowmetry, was reported to have wide limits of agreement, with its accuracy influenced by variables such as smartphone type and toilet settings [81].

Enuresis Alarm Innovations

Novel Configurations

Category Overview and Key Drivers

Enuresis alarms function by detecting urine on 2 conducting strips, causing a voltage change and triggering an alarm [85-87]. Alarms with rigid wires may cause discomfort or psychological effects for the child, whereas fragile sensors can lead to missed alarms due to unintentional or deliberate deactivation. Conversely, overly sensitive alarms may trigger false alarms, which can be discouraging for patients [86]. These design shortcomings related to comfort, durability, and reliability in traditional enuresis alarms are the reason for novel configurations.

Tools Within This Category

We encountered 12% (8/66) of the studies featuring various configurations of enuresis alarms (Table S8 in Multimedia Appendix 3 [40-105]), spanning pad-and-bell to body-worn alarms and wired to wireless alarms. The pad-and-bell alarm used a bed mat with conductive strips, whereas the body-worn alarm had a sensor attached to underwear [86-88]. Wireless versions transmitted signals to an alarm positioned further away, requiring the child to leave the bed to deactivate it [86]. Smart textile sensors integrated conductive threads into underwear or disposable urine pads [85], with recent innovations incorporating wireless transmitters and smartphone apps for notifications and reinforcement strategies [89].

Configurations of the alarm signal itself ranged from loud acoustic alarms to subtler options such as lights, vibrating alarms, or personalized recorded messages [86]. For instance, one study evaluated a code word alarm with a digital voice recorder for parents to record personalized code words, rewarding children for recalling them the following morning [90]. Finally, another, more unconventional approach involved electrical impulses via surface electrode pads to activate the external urethral sphincter to interrupt micturition instead of relying on audible alarms [91,92].

Study Results and Conclusions

Comparative studies on pad-and-bell versus body-worn alarms reported mixed results [87,88]. While one study reported better outcomes with the pad-and-bell alarm [88], another found both equally effective but noted high dropout rates due to embarrassment [87]. Variations in false alarm rates and missed alarms were observed, suggesting that individual device features may have influenced performance more than the alarm type itself.

A review [86] highlighted the potential of smart textiles to enhance comfort and detection efficiency, with studies reporting positive performance and safety during daytime training [85,89]. One of these studies specifically found that a smart textile alarm with an app delivered similar outcomes in less time with high patient and parent satisfaction compared to standard behavioral therapy for toilet training children with autism spectrum disorder [89].

Using a prerecorded code word was reported to improve arousal response and training attitude, reducing dropout by 10%, although it only affected treatment outcomes in a subgroup with monosymptomatic NE [90]. Electrical impulses were reported as effective with perineal placement [91], though based on a small sample, whereas suprapubic placement showed no significant effect [92]. Reported benefits included potentially reducing fully wet beds, boosting self-confidence, and family motivation, but drawbacks such as higher cost, complexity, and potential fear of electrical currents were also noted.

Prevoid Alarms

Category Overview and Key Drivers

Traditional enuresis alarms provide a warning after an enuretic incident has occurred. Consequently, both the child and the parent still experience the inconvenience and consequences of wet pants or sheets during treatment, requiring a substantial level of motivation [93,94] Moreover, the child must establish an indirect link between the full bladder sensation and the alarm as it sounds after the bladder has emptied. These limitations have driven the development of prevoid enuresis alarms, designed to anticipate enuretic incidents by predicting when they might occur and provide a warning in advance.

Tools Within This Category

A total of 20% (13/66) of the studies were found on prevoid alarms (Table S9 in Multimedia Appendix 3 [40-105]), of which most (12/13, 92%) used ultrasound technology to trigger an alarm when bladder volume reached a predefined threshold. Basic ultrasound technology detected the posterior bladder wall rising above the symphysis pubis [93,95], whereas advanced systems estimated bladder volume using an array of transducers to locate both the anterior and posterior bladder walls through ultrasound echoes [94,96-104].

Various specialized measures were implemented to enhance the reliability of prevoid alarms. This included using intelligent software and machine learning to customize for factors such as body morphology, age, and sex [95,98-104]. Furthermore, to mitigate sensitivity to ultrasound probe movement, solutions such as tailored belts [93,95,104], elastic garments [94,98-100], adhesive materials [101-103], and even synchronizing measurements with the child's respiratory rhythm [95] were explored.

One prevoid alarm system estimated bladder volume using wearables and deep learning instead of ultrasound. It tracked impedance using a waist belt sensor, heart rate and sleep stage

XSL•FO
RenderX

using a smartwatch, and limb movement using fabric foot bands [105].

Study Results and Conclusions

The studies on prevoid enuresis alarms were primarily validation and feasibility studies reporting on the technological feasibility of these alarms, with detection rates varying from <50% to 90% [93,95-98,102,104,105]. However, sample sizes were small (1 to 41 patients) or relied on artificial bladder models.

Patient acceptance, influenced by the need for a comfortable, discrete, and nonstigmatizing design, was evaluated for MyPad [99,100] and SENS-U [101,103], both of which were well accepted. SENS-U was also reported to be feasible for both daytime bladder training [101] and nighttime monitoring [103].

Despite feasibility, several challenges were noted, including sensor positioning sensitivity; bladder shape changes due to an enlarged rectum or posture; and compatibility issues with different body types, particularly in children with obesity [93,94,98,100-104]. In addition, daytime observations of SENS-U showed that 23% of children experienced urgency or leakage before reaching the predefined volume-based threshold [101].

Finally, only 8% (1/13) of these studies compared the prevoid alarm with the traditional alarm, reporting comparable cure rates but a treatment duration approximately 3 times longer with the prevoid alarm [94]. The study suggested that awakening to wet sheets may constitute a classic conditioning negative reinforcement potentially essential for the cure process.

Discussion

Principal Findings

In this scoping review, we identified and mapped innovative, technology-driven, digital tools for managing pediatric UI. The field of pediatric UI demonstrates a considerable level of innovation, as evidenced by the inclusion of 66 articles, with nearly one-third (21/66, 32%) focusing on NE. Our analysis led to the identification of six primary categories of tools, 3 of which were divided into subcategories. (1) Digital self-management tools, which leveraged the widespread use of smartphones, tablets, and computers, offering a convenient and accessible medium for health-related purposes such as electronic bladder diaries [40,41] and self-management apps or portals [42-46]. (2) Serious games, which engaged children through playful interactions, such as games for pelvic floor biofeedback training [47-52] and standard urotherapy [53]. (3) Reminder technology, which aimed to enhance therapy adherence without constant parental prompting, such as timer watches [54-56] and self-management reminder tools [57-59]. (4) Educational media for delivering health-related information to patients, which was subdivided into videos and other types of educational media. Videos offered visually engaging, 24/7-accessible, and anonymous dissemination of health information on topics ranging from behavioral therapy, such as bladder training and NE therapy [60-63], to preparing for surgical interventions [64]. A variety of other materials were found that provided child-friendly and engaging patient education, such as UI homework books [65,66], an NE multimedia program [67,68], and educational materials for UTIs [69-71]. (5) Telehealth and RPM for remote health information exchange, which was subdivided into communication and technological advances. Remote communication between patients and health care professionals through videoconferencing [72], home monitoring systems [73], and an embodied conversational agent [74-76] can lead to time, cost, and travel savings. Technological advances such as machine learning algorithms [79], new technologies [78,80,81], and new applications enabling home use [77,78,80-84] aimed to overcome the limitations of in-clinic procedures. (6) Enuresis alarm innovations, which were subdivided into novel configurations of traditional enuresis alarms and prevoid alarms. Enuresis alarm configurations encompassed pad-and-bell and body-worn alarms [86-88], wired and wireless systems [86], smart textiles [85,89], and a variety of alarm signals [86,90-92]. Prevoid alarms aimed to predict enuretic incidents and provide warnings in advance [93-105], with most using bladder ultrasound technology [93-104].

Many of the tools identified in this review were described as promising alternatives to traditional methods, with studies reporting their feasibility [40-46,57-59,65,66,93,95-98,102,104,105], potential to improve outcomes [43,47-53,60,61,64,74-76,82-84,90], and ability to save time for health care providers [49,60,61,64,68-70]. In addition, these tools were described as engaging for children, improving patient compliance [43,46,52,54-56,66,67,74-76] and leading to high levels of patient satisfaction and preference [40-46,53,58,72-77,89,100,101,103].

However, the current body of evidence has methodological limitations, including retrospective designs [49,55,82,83]; absence of control groups [47,57,75,76,84,94,101,103]; and small sample sizes, with almost one-quarter of the studies (16/66, 24%) having a sample size of ≤15 participants [45,58,59,73,75,83,85,91,96,98-101,103-105]. While 21% (14/66) of the included studies were RCTs [43,50-52,56,60,61,64,68,77,88-90,92], some of these (3/14, 21%) did not compare equivalent therapies [43,50,51]. Furthermore, the findings varied, with some studies (10/66, 15%) demonstrating differences in outcomes between the novel tools and traditional methods [49,55,56,74-76,88-90,94], whereas others (9/66, 14%) reported similar outcomes [40,50,60,68-70,77,87,92]. These observations highlight the need for more rigorous research to better establish the effectiveness of these tools.

Understanding the Challenges of Managing Pediatric UI

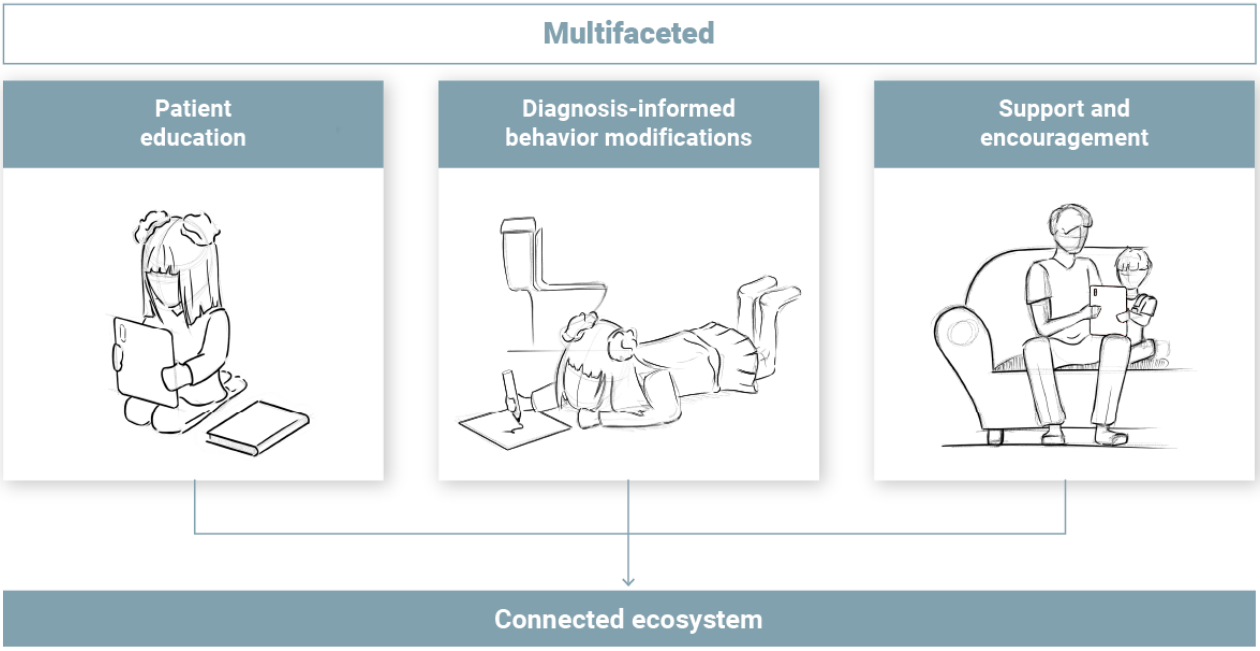
The variability in the reported findings and gaps in robust evidence may also reflect the complexity of managing pediatric UI. This review highlights the multifaceted and multifactorial influences on pediatric UI management.

Multifaceted Influences

Pediatric UI management encompasses various aspects, such as patient education, diagnosis-informed behavior modifications, and providing support and encouragement (Figure 3).

XSL•FO
RenderX

Figure 3. Pediatric urinary incontinence management is multifaceted, including patient education, diagnosis-informed behavior modifications, and support and encouragement. Integrating these aspects into a connected ecosystem might be an ideal way to address this multifaceted nature.



Regarding patient education, the studies included in this review highlighted various educational activities, such as raising awareness [69-71], explaining complex concepts [64], setting expectations [67,68], providing lifestyle advice [60-63,75,76], and promoting adherence [53,65,66,74].

Regarding diagnosis-informed behavior modifications, management typically began with a thorough diagnostic process, as described in several included studies (9/66, 14%) [40,65,72,74-76,78,80,81]. By accurately identifying contributing factors and underlying causes, appropriate therapeutic interventions can be selected. The interventions described in the reviewed studies included providing instructions and reminders to establish healthy bladder and bowel habits [54-59]. Monitoring these habits through the collection of data on fluid intake, voiding patterns, and bowel movements [40,41,53], as well as quality of life indicators [42-45], was reported to help identify patterns and track progress over time [42-46,53,73].

Regarding support and encouragement, the included studies emphasized that support, encouragement, and feedback can help reinforce positive behaviors and maintain therapy consistency whether in the clinic [47-52], at home [42-45,72,74-77,84], or in a group setting [82].

Despite this multifaceted nature, many tools discussed in this review offered stand-alone solutions that addressed specific aspects. Examples include mobile health apps for keeping electronic bladder diaries [40,41], biogames for pelvic floor biofeedback [47-52], timer watches for timed voiding [54-56], educational videos covering specific concepts [60-64], and enuresis alarm innovations for alarm training [85-104]. While these stand-alone solutions addressed particular aspects of pediatric UI management, they did so in isolation, potentially missing opportunities to create synergies and links between different aspects of UI management. This review leads us to wonder whether integrating these various aspects into a connected ecosystem could better address the multifaceted nature of UI management.

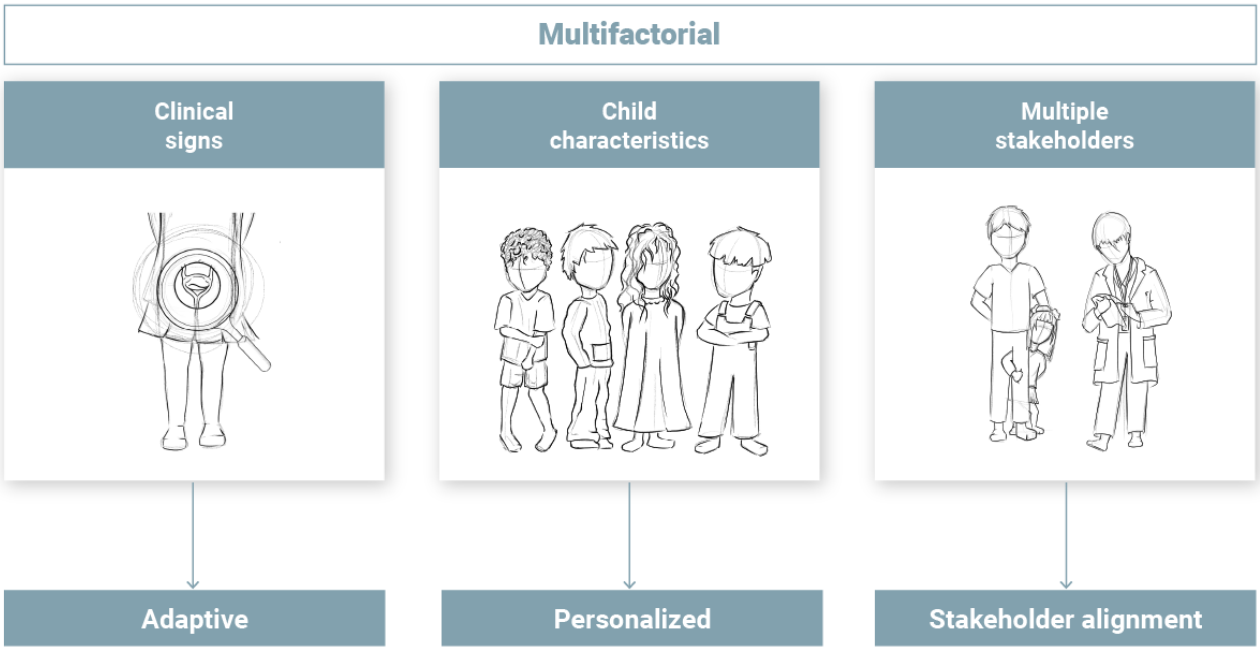
Multifactorial Influences

Pediatric UI management is influenced by a range of factors, such as clinical signs, child characteristics, and the involvement of multiple stakeholders (Figure 4).

Regarding clinical signs, the management of pediatric UI depends on the specific type and severity of symptoms that children experience. These can include multiple symptoms related to the bladder and bowel [40,60,61,72,78,80,81] occurring during the day or night [41,74-76,79]. Therefore, each child's therapy journey is unique, with varying underlying causes, progress rates, and potential setbacks. Accurate differentiation during the diagnostic phase, along with continuous monitoring of therapy progress, as described in some studies (11/66, 17%) [42-45,53,74-77,82,83], can support tailoring management plans to the child's specific condition and adapting them as needed. We propose that tools be adaptive to meet the evolving needs of children during therapy and help maintain their motivation. For instance, feedback from children using the urotherapy serious game indicated a desire for more dynamic gameplay that adapts to their training progress [53].

XSL•FO
RenderX

Figure 4. Pediatric urinary incontinence management is multifactorial, influenced by a range of factors such as (1) clinical signs that progress over time, requiring adaptive interventions; (2) child characteristics that demand a personalized approach; and (3) the involvement of multiple stakeholders whose needs must be aligned.



Regarding child characteristics, the management of UI depends not only on the child's clinical signs and physiological needs but also on their psychological and emotional needs. As described in the studies, these needs can influence the child's motivation [52,53,93,94,106], attitude toward training [43,46,55,56,65,66,74,90], and level of maturation and attention span [52,67]. Emotional factors such as self-image, embarrassment, fear, and stress were also described as key influences [59,64,87,91]. These child-specific characteristics can influence outcomes. For instance, varying findings of comparative studies on different enuresis alarm configurations [87,88] suggested that certain configurations may be more effective for some children than for others. In addition, one study reported that reminder technology was an equally effective but simpler alternative to enuresis alarms in treating DUI [54]. We propose that personalizing or selecting tools based on the child's preferences may enhance acceptance and satisfaction and potentially lead to better outcomes.

Regarding multiple stakeholders, each stakeholder involved in pediatric UI management has distinct needs. The studies indicated that parental motivation often depended on constraints such as time, cost, and travel, prioritizing convenience and self-efficacy [22,25,40,41,46,72-76,78,89,91,93,94]. Health care professionals, on the other hand, were described as managing time constraints in busy clinical practices, aiming for efficiency and high-quality data [40,46,49,60,61,64,78,80] while also maintaining empathy and establishing a therapeutic alliance [72,75]. Aligning these diverse needs can be challenging, as described for the UTI remote monitoring system [73]. Although the system comprised a set of products, most parents primarily used the urinalysis device because it met their needs for self-efficacy and reassurance. The other products were underused, leading to frustration among health care providers, who faced added workload for setting up the system without improved data quality. These findings underscore the importance of designing tools that balance the needs of all stakeholders, improving quality of life for patients and families while optimizing time and efficiency for health care professionals.

Transforming Health Care With Connected Ecosystems

Given the multifaceted and multifactorial nature of pediatric UI management, we propose that novel tools should be designed as health care ecosystems that are (1) connected, (2) adaptive, (3) personalized, and (4) aligned with the needs of all stakeholders.

The concept of health care ecosystems for managing pediatric UI is rooted in a broader transformation of the health care industry driven by innovation [115]. Advances such as electronic health records, networked computer systems, and telehealth have revolutionized health care delivery, shifting from a model centered on face-to-face interactions with individual clinicians to a smart, interconnected health care community emphasizing shared decision-making [116]. The COVID-19 pandemic further accelerated this transition, highlighting the critical role of digital technologies in maintaining access to health care [117]. Furthermore, recent breakthroughs in computational power and artificial intelligence promise to expand the capabilities of health care systems even further [116].

A fair part of the currently available medical devices already has the ability to connect and communicate with other devices or systems. This connectivity transforms previously isolated tools into collaborative systems that operate within closed feedback loops, tailoring outputs to the individual [118]. For example, in glucose monitoring, sensors measure blood glucose levels, which then inform insulin delivery systems to maintain optimal glucose control [119]. Similarly, smart pillboxes

XSL•FO
RenderX

improve medication adherence by tracking remaining doses, providing reminders through a connected alarm device [120].

While many of these systems have been tested primarily in contexts such as monitoring older patients at home [121]—where they support independence and reduce health care burdens—pediatric urology presents a significant opportunity for similar innovations. For example, a recently published article on the Minze Homeflow system demonstrates how a connected ecosystem can enhance pediatric UI management [122]. It integrates physical products (uroflowmeter and accessories) with digital tools (Minze app and clinician portal) for remote monitoring of uroflowmetry and bladder diary data. By shifting diagnostics to the home environment, the Homeflow system minimizes stress for pediatric patients while providing more reliable data compared to in-clinic measurements. Such a system demonstrates the feasibility and benefits of connected health care ecosystems in addressing specific challenges in pediatric urology.

Another study demonstrated how the Homeflow ecosystem could be further expanded by connecting additional products to the system, such as a smart drinking bottle and reminder technology, to personalize urotherapy [123]. The smart drinking bottle and home uroflowmeter monitored fluid intake and urinary output, whereas a reminder watch linked these devices, offering personalized voiding prompts based on the child's bladder capacity and fluid intake. This demonstrates the potential of connected ecosystems to adapt and align with the individual care needs of pediatric patients.

Overcoming Challenges in Health Care Innovation

The successful implementation of health care innovation and ecosystems requires addressing important challenges, such as regulatory and privacy concerns, equitable access, screen time considerations, and keeping health care human centered. Thoughtful strategies are needed to overcome these challenges and maximize benefits.

Regulatory and Privacy Concerns

Ensuring patient privacy and data security remains a critical challenge for digital health care innovation [124]. As these tools handle sensitive patient data, robust regulatory frameworks must be established to protect patient data. However, balancing innovation with regulation remains a challenge, especially considering the rapid pace of technological advancements. Current regulations may struggle to keep up, potentially delaying the adoption of transformative tools. A balanced approach is required to ensure safety and trust without hampering innovation that can improve patient care.

Equity in Access

Ensuring equitable access is another challenge for health care innovation, especially in low-resource settings. Individuals in such settings must not be left behind as they are often the ones who stand to benefit the most from these advancements [115]. However, barriers such as limited infrastructure, potentially high entry costs, and digital literacy disparities can hinder adoption [116]. Nevertheless, innovation also has the potential to improve equity by decentralizing care. Shifting diagnostics and therapies from clinics to homes or communities can reduce costs and improve accessibility. In addition, customizable solutions tailored to varying resource levels, aligned with reimbursement models, or supported through public health initiatives can support the equitable implementation of innovation.

Screen Time Considerations

The use of digital tools raises concerns about screen time and its potential impact on children's health. A 2019 systematic review [125] identified associations between excessive screen time and health risks, including adiposity, unhealthy diet, depressive symptoms, and reduced quality of life. However, the review did not consider the influence of content or context on these outcomes. Research indicates that well-designed, age-appropriate content, especially when curated by engaged adults [126], can positively impact learning and behavior [127]. For example, screen time can calm distressed children during medical procedures [128] or support language learning [129]. Moreover, active video games or nature exploration apps can encourage physical activity and imaginative play by connecting on- and off-screen experiences [130].

Nevertheless, certain screen use patterns present clear risks. Screen time before bedtime is strongly associated with sleep disturbances [131]. The mere presence of electronic devices in bedrooms has been linked to shorter sleep duration, partly due to melatonin suppression [132]—an essential consideration for tools addressing NE. A 2024 systematic review [133] reinforced these findings, showing that greater use of mobile devices, particularly at bedtime or in patterns of device dependence, can harm mental health in children and adolescents. At the same time, it emphasized technology's dual impact—while social media often correlates with negative effects, other communication tools can enhance psychological well-being by strengthening social connections and providing access to support. These effects vary across individuals, highlighting the need to identify those most vulnerable to adverse effects.

In summary, the impact of screen time is shaped by multiple factors. To maximize benefits, digital tools should be thoughtfully designed to deliver engaging, age-appropriate content; promote balanced screen habits; integrate on- and off-screen interactions; and avoid bedtime use and overreliance on devices.

Keeping Health Care Human Centered

Digital technology presents significant opportunities to enhance efficiency and accuracy in health care. However, human interaction remains fundamental to preserving empathy, trust, and nuanced clinical decision-making. For example, a study in this review reported that health care providers experienced difficulties building rapport with pediatric patients due to technical barriers [72]. This underscores the importance of designing tools that augment health care professionals' capabilities, ensuring that technology supports rather than depersonalizes patient care. Therefore, a human-centered approach to technology is needed to actively incorporate the perspectives and needs of patients, caregivers, and health care

XSL•FO
RenderX

providers, placing them at the center of smart and connected health care in both research and practice [116].

Such an approach values not only objective evidence but also subjective patient and stakeholder experiences, combining quantitative and qualitative methods. However, in our review, only 12% (8/66) of the studies used qualitative [42,72,99] or mixed methods [40,44,58,67,73] designs. Despite this, more than one-third (25/66, 38%) of studies evaluated outcomes related to patient experience, such as satisfaction, motivation, and usability [42-45,52,53,57,58,66,67,71-73,75,76,82,88-90,94,98-101,103]. This indicates a growing recognition of the importance of patient-centered outcomes but also underscores the need for more qualitative approaches to capture deeper insights.

To incorporate a human-centered approach, we conducted a separate qualitative study alongside this scoping review to explore children's perspectives, experiences, and expectations regarding urotherapy-supporting tools. To facilitate this, we developed a focus group discussion toolkit specifically tailored to pediatric UI [134]. The toolkit incorporated creative tasks involving imagination and play, enabling children to express their thoughts and experiences in a child-friendly and engaging manner. By combining insights from this scoping review with the findings of the focus groups, we aimed to provide a comprehensive perspective that blends existing evidence with the desires and needs of the target population. This integrated approach can guide the development of more effective and user-centered tools for pediatric UI management.

Limitations

This scoping review has certain methodological limitations that should be considered when interpreting its results. First, the search was limited to 3 electronic databases, and only English-language papers were included. Second, only 1 reviewer screened the entire set of search results. However, the team engaged in frequent and extensive discussions throughout the study selection process to ensure rigorous screening. Third, our search had an extensive scope due to our broad criteria for defining innovative, technology-driven, digital tools, as well as having no time limits. The broad field and the potential abundance of relevant tools made it challenging to provide a concise overview of all the information. Consequently, this review encompassed a broad range of studies and a wide variety of innovative, technology-driven, digital tools.

Conclusions

This scoping review offers valuable insights into various categories of innovative, technology-driven, digital tools for managing pediatric UI and their supporting evidence. The inclusion of 66 articles highlights a considerable level of innovation in the field.

We identified 6 main categories of tools: digital self-management, serious games, reminder technology, educational media, telehealth and RPM, and enuresis alarm innovations. Further subcategories were found within these main categories, such as video and other media within educational media, communication and technological advances within telehealth and RPM, and novel configurations and prevoid alarms within enuresis alarm innovations.

Many of these tools were described as promising alternatives to traditional methods, with studies reporting feasibility, effectiveness in improving outcomes, and potential time savings for health care providers. In addition, these tools were described as engaging children effectively, potentially enhancing patient compliance and contributing to high patient satisfaction and preference. However, this review identified gaps in research, highlighting the need for more rigorous research to better assess the effectiveness of these tools and address the complex, multifaceted challenges of pediatric UI management. We propose the development of connected, adaptive, and personalized health care ecosystems that integrate diverse strategies for UI management, including education, diagnosis, therapy, monitoring, and feedback. By enabling synergies and links between these different strategies, such ecosystems could create a closed-loop system that enhances the alignment and efficiency of care.

We suggest that future research and development in this field adopt a human-centered and multidisciplinary approach, combining quantitative and qualitative methods to incorporate user perspectives and involve all stakeholders from the start. This mixed methods approach may help ensure that both objective evidence and subjective experiences are equally valued, supporting impactful innovations in the complex field of pediatric UI management.

Acknowledgments

The generative artificial intelligence tool ChatGPT by OpenAI was used to improve language and readability. The tool provided inspiration on how to rephrase parts of the manuscript, but final edits and decisions were made by the authors. This research is part of a PhD project partially funded by the Flanders Innovation and Entrepreneurship Baekeland grant (project ID HBC.2021.0836) and the company Minze Health.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

The conceptualization of this scoping review, including the formulation of the research question, search strategy, and eligibility criteria, was a collaborative effort involving the entire team of authors. LB took the lead in the study selection process and data

XSL•FO
RenderX

analysis, which included screening titles and abstracts, assessing full-text articles, charting the data, and conducting categorization of the identified tools. Throughout these processes, LB consulted with the team to ensure consistency, rigor, and consensus. GDW and AV provided expertise on pediatric urinary incontinence management, whereas LVC and JV contributed insights on product design and categorization of the tools. All authors offered guidance throughout the research project, including interpretation and critical review. LB drafted the manuscript and finalized the version presented. The final manuscript was critically reviewed and approved by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR checklist.
[DOCX File , 108 KB - ijmr_v14i1e66336_app1.docx]

Multimedia Appendix 2
Full search strategy.
[DOCX File , 14 KB - ijmr_v14i1e66336_app2.docx]

Multimedia Appendix 3
Overview of the included studies by tool category.
[DOCX File , 1060 KB - ijmr_v14i1e66336_app3.docx]

References

1. Nieuwhof-Leppink AJ, Schroeder RP, van de Putte EM, de Jong TP, Schappin R. Daytime urinary incontinence in children and adolescents. Lancet Child Adolesc Health 2019 Jul;3(7):492-501. [doi: 10.1016/s2352-4642(19)30113-0]
2. Spee-van der Wekke J, Hirasing RA, Meulmeester JF, Radder JJ. Childhood nocturnal enuresis in The Netherlands. Urology 1998 Jun;51(6):1022-1026. [doi: 10.1016/s0090-4295(98)00106-x] [Medline: 9609644]
3. Marugán de Miguelsanz JM, Lapeña López de Armentia S, Rodríguez Fernández LM, Palau Benavides MT, Torres Hinojal MC, Menau Martín G, et al. [An epidemiological analysis of the sequence of bladder control and nocturnal enuresis prevalence in the children of the province of León]. An Esp Pediatr 1996 Jun;44(6):561-567. [Medline: 8849099]
4. Lottmann HB. [Treatment of nocturnal enuresis in France]. Presse Med 2000 May 20;29(18):987-990. [Medline: 10862246]
5. Thibodeau BA, Metcalfe P, Koop P, Moore K. Urinary incontinence and quality of life in children. J Pediatr Urol 2013 Feb;9(1):78-83. [doi: 10.1016/j.jpurol.2011.12.005] [Medline: 22236468]
6. Bower W. Self-reported effect of childhood incontinence on quality of life. J Wound Ostomy Continence Nurs 2008;35(6):617-621. [doi: 10.1097/01.WON.0000341476.71685.78] [Medline: 19018203]
7. Butler RJ. Annotation: night wetting in children: psychological aspects. J Child Psychol Psychiatry 2003 Oct 09;39(4):453-463. [doi: 10.1111/1469-7610.00342]
8. Morison MJ, Tappin D, Staines H. 'You feel helpless, that's exactly it': parents' and young people's control beliefs about bed-wetting and the implications for practice. J Adv Nurs 2000 May 25;31(5):1216-1227. [doi: 10.1046/j.1365-2648.2000.01426.x] [Medline: 10840256]
9. Deshpande AV, Craig JC, Smith GH, Caldwell PH. Factors influencing quality of life in children with urinary incontinence. J Urol 2011 Sep;186(3):1048-1052. [doi: 10.1016/j.juro.2011.04.104] [Medline: 21784481]
10. Austin PF, Bauer SB, Bower W, Chase J, Franco I, Hoebeke P, et al. The standardization of terminology of lower urinary tract function in children and adolescents: update report from the standardization committee of the International Children's Continence Society. Neurourol Urodyn 2016 Apr;35(4):471-481. [doi: 10.1002/nau.22751] [Medline: 25772695]
11. dos Santos J, Varghese A, Williams K, Koyle MA. Recommendations for the management of bladder bowel dysfunction in children. Pediat Therapeut 2014;04:191. [doi: 10.4172/2161-0665.1000191]
12. Santos J, Lopes RI, Koyle MA. Bladder and bowel dysfunction in children: an update on the diagnosis and treatment of a common, but underdiagnosed pediatric problem. Can Urol Assoc J 2017 Feb 16;11(1-2Suppl1):S64-S72 [FREE Full text] [doi: 10.5489/cuaj.4411] [Medline: 28265323]
13. Hellström AL, Hjälmås K, Jodal U. Rehabilitation of the dysfunctional bladder in children: method and 3-year followup. J Urol 1987 Oct;138(4):847-849. [doi: 10.1016/s0022-5347(17)43395-7] [Medline: 3656544]
14. Nevéus T, von Gontard A, Hoebeke P, Hjälmås K, Bauer S, Bower W, et al. The standardization of terminology of lower urinary tract function in children and adolescents: report from the standardisation committee of the International Children's Continence Society. J Urol 2006 Jul;176(1):314-324. [doi: 10.1016/s0022-5347(06)00305-3]
15. Chang SJ, Van Laecke E, Bauer SB, von Gontard A, Bagli D, Bower WF, et al. Treatment of daytime urinary incontinence: a standardization document from the International Children's Continence Society. Neurourol Urodyn 2017 Jan 16;36(1):43-50. [doi: 10.1002/nau.22911] [Medline: 26473630]

XSL•FO
RenderX

16. Nieuwhof-Leppink AJ, Hussong J, Chase J, Larsson J, Renson C, Hoebeke P, et al. Definitions, indications and practice of urotherapy in children and adolescents: - a standardization document of the International Children's Continence Society (ICCS). J Pediatr Urol 2021 Apr;17(2):172-181. [doi: 10.1016/j.jpurol.2020.11.006] [Medline: 33478902]
17. van Gool JD, de Jong TP, Winkler-Seinstra P, Tamminen-Möbius T, Lax H, Hirche H, et al. Multi-center randomized controlled trial of cognitive treatment, placebo, oxybutynin, bladder training, and pelvic floor training in children with functional urinary incontinence. Neurourol Urodyn 2014 Jun 15;33(5):482-487. [doi: 10.1002/nau.22446] [Medline: 23775924]
18. Schäfer SK, Niemczyk J, von Gontard A, Pospeschill M, Becker N, Equit M. Standard urotherapy as first-line intervention for daytime incontinence: a meta-analysis. Eur Child Adolesc Psychiatry 2018 Aug;27(8):949-964. [doi: 10.1007/s00787-017-1051-6] [Medline: 28948380]
19. Vijverberg MA, Stortelder E, de Kort LM, Kok ET, de Jong TP. Long-term follow-up of incontinence and urge complaints after intensive urotherapy in childhood (75 patients followed up for 16.2-21.8 years). Urology 2011 Dec;78(6):1391-1396. [doi: 10.1016/j.urology.2011.08.055] [Medline: 22014960]
20. Nevéus T, Fonseca E, Franco I, Kawauchi A, Kovacevic L, Nieuwhof-Leppink A, et al. Management and treatment of nocturnal enuresis-an updated standardization document from the International Children's Continence Society. J Pediatr Urol 2020 Feb;16(1):10-19. [doi: 10.1016/j.jpurol.2019.12.020] [Medline: 32278657]
21. Glazener CM, Evans JH, Peto RE. Alarm interventions for nocturnal enuresis in children. Cochrane Database Syst Rev 2003(2):CD002911. [doi: 10.1002/14651858.CD002911] [Medline: 12804443]
22. Nock MK, Photos V. Parent motivation to participate in treatment: assessment and prediction of subsequent participation. J Child Fam Stud 2006 May 3;15(3):333-346. [doi: 10.1007/s10826-006-9022-4]
23. Mulders MM, Cobussen-Boekhorst H, de Gier RP, Feitz WF, Kortmann BB. Urotherapy in children: quantitative measurements of daytime urinary incontinence before and after treatment according to the new definitions of the International Children's Continence Society. J Pediatr Urol 2011 Apr;7(2):213-218. [doi: 10.1016/j.jpurol.2010.03.010] [Medline: 20541978]
24. Hyuga T, Nakamura S, Kawai S, Nakai H. Evaluation of the effectiveness of a short-term treatment and repeat treatment of nocturnal enuresis using an enuresis alarm. Urology 2017 Jul;105:153-156. [doi: 10.1016/j.urology.2017.01.005] [Medline: 28089888]
25. Oğuz U, Sarıkaya S, Özyuvalı E, Şenocak Ç, Halis F, Çiftci M, et al. Family compliance with the use of alarm devices in the treatment of monosymptomatic nocturnal enuresis. Turk J Urol 2014 Mar 24;40(1):52-55 [FREE Full text] [doi: 10.5152/tud.2014.35033] [Medline: 26328146]
26. Baker S, Richards D, Caldwell P. Relational agents to promote eHealth advice adherence. In: Proceedings of the 13th Pacific Rim International Conference on Artificial Intelligence. 2014 Presented at: PRICAI 2014; December 1-5, 2014; Gold Coast, Australia. [doi: 10.1007/978-3-319-13560-1_87]
27. Barbosa CD, Balp MM, Kulich K, Germain N, Rofail D. A literature review to explore the link between treatment satisfaction and adherence, compliance, and persistence. Patient Prefer Adherence 2012;6:39-48 [FREE Full text] [doi: 10.2147/PPA.S24752] [Medline: 22272068]
28. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [doi: 10.1016/S0197-2456(02)00320-3]
29. Tincello DG, Williams KS, Joshi M, Assassa RP, Abrams KR. Urinary diaries: a comparison of data collected for three days versus seven days. Obstet Gynecol 2007;109(2 Part 1):277-280. [doi: 10.1097/01.aog.0000252832.21986.c8]
30. Mangera A, Marzo A, Heron N, Fernando D, Hameed K, Soliman AH, et al. Development of two electronic bladder diaries: a patient and healthcare professionals pilot study. Neurourol Urodyn 2014 Sep 02;33(7):1101-1109. [doi: 10.1002/nau.22469] [Medline: 24000163]
31. Apos E, Schuster S, Reece J, Whitaker S, Murphy K, Golder J, et al. Enuresis management in children: retrospective clinical audit of 2861 cases treated with practitioner-assisted bell-and-pad alarm. J Pediatr 2018 Feb;193:211-216. [doi: 10.1016/j.jpeds.2017.09.086] [Medline: 29246468]
32. Joiner R, Gavin J, Brosnan M, Cromby J, Gregory H, Guiller J, et al. Comparing first and second generation digital natives' internet use, internet anxiety, and internet identification. Cyberpsychol Behav Soc Netw 2013 Jul;16(7):549-552. [doi: 10.1089/cyber.2012.0526] [Medline: 23675995]
33. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. Pain 2004 Feb;107(3):213-219. [doi: 10.1016/j.pain.2003.10.005] [Medline: 14736583]
34. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RS, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. Lancet Digit Health 2020 Apr;2(4):e168-e178. [doi: 10.1016/s2589-7500(20)30017-0]
35. Heerman WJ, Rothman RL, Sanders LM, Schildcrout JS, Flower KB, Delamater AM, Greenlight Investigators, et al. A digital health behavior intervention to prevent childhood obesity: the greenlight plus randomized clinical trial. JAMA 2024 Dec 24;332(24):2068-2080. [doi: 10.1001/jama.2024.22362] [Medline: 39489149]

XSL•FO
RenderX

36. Ramsey RR, Plevinsky JM, Kollin SR, Gibler RC, Guilbert TW, Hommel KA. Systematic review of digital interventions for pediatric asthma management. J Allergy Clin Immunol Pract 2020 Apr;8(4):1284-1293 [FREE Full text] [doi: 10.1016/j.jaip.2019.12.013] [Medline: 31870809]
37. 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 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
38. Niksch AL. mHealth in pediatrics-finding healthcare solutions for the next generation. Mhealth 2015;1:7 [FREE Full text] [doi: 10.3978/j.issn.2306-9740.2015.03.09] [Medline: 28293567]
39. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May 15;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
40. Johnson EK, Estrada CR, Johnson KL, Nguyen HT, Rosoklija I, Nelson CP. Evaluation of a mobile voiding diary for pediatric patients with voiding dysfunction: a prospective comparative study. J Urol 2014 Sep;192(3):908-913 [FREE Full text] [doi: 10.1016/j.juro.2014.03.099] [Medline: 24704008]
41. Myint M, Adam A, Herath S, Smith G. Mobile phone applications in management of enuresis: the good, the bad, and the unreliable!. J Pediatr Urol 2016 Apr;12(2):112.e1-112.e6. [doi: 10.1016/j.jpurol.2015.09.011] [Medline: 26611873]
42. Whale K, Beasant L, Wright AJ, Yardley L, Wallace LM, Moody L, et al. A smartphone app for supporting the self-management of daytime urinary incontinence in adolescents: development and formative evaluation study of URApp. JMIR Pediatr Parent 2021 Nov 15;4(4):e26212 [FREE Full text] [doi: 10.2196/26212] [Medline: 34779780]
43. Bonnert M, Olén O, Lalouni M, Benninga MA, Bottai M, Engelbrektsson J, et al. Internet-delivered cognitive behavior therapy for adolescents with irritable bowel syndrome: a randomized controlled trial. Am J Gastroenterol 2017 Jan;112(1):152-162 [FREE Full text] [doi: 10.1038/ajg.2016.503] [Medline: 27845338]
44. Choi EK, Jung E, Bae E, Ji Y, Lee A. Two-step integrative education program and mHealth for Korean children with spina bifida: a quasi-experimental pre-post study. J Pediatr Nurs 2020 Mar;51:e92-e99. [doi: 10.1016/j.pedn.2019.12.006] [Medline: 31924471]
45. Choi EK, Jung E, Ji Y, Bae E. A 2-step integrative education program and mHealth for self-management in Korean children with spina bifida: feasibility study. J Pediatr Nurs 2019 Nov;49:e54-e62. [doi: 10.1016/j.pedn.2019.09.002] [Medline: 31519400]
46. Fine AM, Kalish LA, Forbes P, Goldmann D, Mandl KD, Porter SC. Parent-driven technology for decision support in pediatric emergency care. Jt Comm J Qual Patient Saf 2009 Jun;35(6):307-315. [doi: 10.1016/s1553-7250(09)35044-8] [Medline: 19565690]
47. McKenna PH, Herndon CD, Connery S, Ferrer FA. Pelvic floor muscle retraining for pediatric voiding dysfunction using interactive computer games. J Urol 1999 Sep;162(3 Part 2):1056-1062. [doi: 10.1016/s0022-5347(01)68065-0]
48. Herndon CD, DeCambre M, McKenna PH. Changing concepts concerning the management of vesicoureteral reflux. J Urol 2001 Oct;166(4):1439-1443. [Medline: 11547107]
49. Kaye JD, Palmer LS. Animated biofeedback yields more rapid results than nonanimated biofeedback in the treatment of dysfunctional voiding in girls. J Urol 2008 Jul;180(1):300-305. [doi: 10.1016/j.juro.2008.03.078] [Medline: 18499170]
50. Kajbafzadeh AM, Sharifi-Rad L, Ghahestani SM, Ahmadi H, Kajbafzadeh M, Mahboubi AH. Animated biofeedback: an ideal treatment for children with dysfunctional elimination syndrome. J Urol 2011 Dec;186(6):2379-2384. [doi: 10.1016/j.juro.2011.07.118] [Medline: 22019033]
51. Ladi-Seyedian S, Kajbafzadeh AM, Sharifi-Rad L, Shadgan B, Fan E. Management of non-neuropathic underactive bladder in children with voiding dysfunction by animated biofeedback: a randomized clinical trial. Urology 2015 Jan;85(1):205-210. [doi: 10.1016/j.urology.2014.09.025] [Medline: 25444633]
52. Oktar T, Dönmez M, Özkuvancı Ü, Ander H, Ziylan O. Animated versus non-animated biofeedback therapy for dysfunctional voiding treatment: does it change the outcome? J Pediatr Surg 2018 Apr;53(4):825-827. [doi: 10.1016/j.jpedsurg.2017.06.002] [Medline: 28648880]
53. Nieuwhof-Leppink AJ, de Jong TP, van de Putte EM, Schappin R. Does a serious game increase intrinsic motivation in children receiving urotherapy? J Pediatr Urol 2019 Feb;15(1):36.e1-36.e7. [doi: 10.1016/j.jpurol.2018.09.003] [Medline: 30401600]
54. Halliday S, Meadow SR, Berg I. Successful management of daytime enuresis using alarm procedures: a randomly controlled trial. Arch Dis Child 1987 Feb 01;62(2):132-137 [FREE Full text] [doi: 10.1136/adc.62.2.132] [Medline: 3548603]
55. Hagstroem S, Rittig N, Kamperis K, Mikkelsen MM, Rittig S, Djurhuus JC. Treatment outcome of day-time urinary incontinence in children. Scand J Urol Nephrol 2008 Jul 09;42(6):528-533. [doi: 10.1080/00365590802098367] [Medline: 18609267]
56. Hagstroem S, Rittig S, Kamperis K, Djurhuus JC. Timer watch assisted urotherapy in children: a randomized controlled trial. J Urol 2010 Oct;184(4):1482-1488. [doi: 10.1016/j.juro.2010.06.024] [Medline: 20727552]
57. Bartos NS, Hamilton S, Barnes E, Caldwell PH. Toilet training children with special needs using a Potty Monkey toy device. J Paediatr Child Health 2021 Mar 19;57(3):365-370. [doi: 10.1111/jpc.15222] [Medline: 33078489]

XSL•FO
RenderX

58. Jennings M, Guilfoyle A, Green J, Cleary Y, Gowran RJ. Octopus watch fosters family resilience by enhancing occupational engagement for children with spina bifida and/or hydrocephalus: pilot study. Int J Environ Res Public Health 2020 Nov 10;17(22):8316 [FREE Full text] [doi: 10.3390/ijerph17228316] [Medline: 33182784]
59. Flannery MA, Butterbaugh GJ, Rice DA, Rice JC. Reminding technology for prospective memory disability: a case study. Pediatr Rehabil 1997 Aug 05;1(4):239-244. [doi: 10.3109/17518429709167365] [Medline: 9689261]
60. Braga LH, Rickard M, Farrokhyar F, Jegatheeswaran K, Brownrigg N, Li C, et al. Bladder training video versus standard urotherapy for bladder and bowel dysfunction: a noninferiority randomized, controlled trial. J Urol 2017 Mar;197(3 Pt 2):877-884. [doi: 10.1016/j.juro.2016.08.089] [Medline: 27569433]
61. Brownrigg N, Braga LH, Rickard M, Farrokhyar F, Easterbrook B, Dekirmendjian A, et al. The impact of a bladder training video versus standard urotherapy on quality of life of children with bladder and bowel dysfunction: a randomized controlled trial. J Pediatr Urol 2017 Aug;13(4):374.e1-374.e8. [doi: 10.1016/j.jpurol.2017.06.005] [Medline: 28733159]
62. Toprak T, Tokat E. A quality analysis of nocturnal enuresis videos on YouTube. J Pediatr Urol 2021 Aug;17(4):449.e1-449.e6. [doi: 10.1016/j.jpurol.2021.03.014] [Medline: 33824069]
63. Nishizaki N, Hirano D, Oishi K, Shimizu T. YouTube videos in Japanese as a source of information on nocturnal enuresis: a content-quality and reliability analysis. Pediatr Int 2022 Jan 05;64(1):e15049. [doi: 10.1111/ped.15049] [Medline: 34747553]
64. Batuman A, Gulec E, Turktan M, Gunes Y, Ozcengiz D. Preoperative informational video reduces preoperative anxiety and postoperative negative behavioral changes in children. Minerva Anestesiol 2016 May;82(5):534-542. [Medline: 26474268]
65. Bower WF, Moore KH, Adams RD. A novel clinical evaluation of childhood incontinence and urinary urgency. J Urol 2001 Dec;166(6):2411-2415. [doi: 10.1016/s0022-5347(05)65604-2]
66. Rhodes C, Timmons S. A feasibility study of an educational intervention for children with daytime wetting. Paediatr Nurs 2008 Jun;20(5):38-42. [doi: 10.7748/paed2008.06.20.5.38.c8261] [Medline: 18816917]
67. Evans JH, Collier J, Crook I, Garrud P, Harris P, MacKinlay DR, et al. Using multimedia for patient information--a program about nocturnal enuresis. Br J Urol 1998 May 04;81 Suppl 3(s3):120-122. [doi: 10.1046/j.1464-410x.1998.00023.x] [Medline: 9634036]
68. Redsell SA, Collier J, Garrud P, Evans JHC, Cawood C. Multimedia versus written information for nocturnal enuresis education: a cluster randomized controlled trial. Child Care Health Dev 2003 Mar 11;29(2):121-129. [doi: 10.1046/j.1365-2214.2003.00321.x] [Medline: 12603357]
69. Teo S, Cheek JA, Craig S. Improving clean-catch contamination rates: a prospective interventional cohort study. Emerg Med Australas 2016 Dec 24;28(6):698-703. [doi: 10.1111/1742-6723.12697] [Medline: 27778452]
70. Jacob R, John-Denny B, Donaldson K, White B, Outhred A, McCaskill ME. Educational intervention does not reduce non-invasive urine contamination rates in children presenting to the emergency department. J Paediatr Child Health 2020 Jan 19;56(1):142-147. [doi: 10.1111/jpc.14542] [Medline: 31216105]
71. Kaur T, Dahiya S, Satija SH, Nawal SJ, Kshetrimayum N, Ningthoujam J, et al. Foldscope as a primary diagnostic tool for oral and urinary tract infections and its effectiveness in oral health education. J Microsc 2020 Jul;279(1):39-51. [doi: 10.1111/jmi.12896] [Medline: 32286690]
72. Kuo AC, Geleris B, Arnhym A. TeleBBD: telehealth management of bladder bowel dysfunction - a provider survey of benefits and limitations. J Pediatr Urol 2021 Dec;17(6):792.e1-792.e7. [doi: 10.1016/j.jpurol.2021.09.021] [Medline: 34656434]
73. Carter B, Whittaker K, Sanders C. Evaluating a telehealth intervention for urinalysis monitoring in children with neurogenic bladder. J Child Health Care 2019 Mar 27;23(1):45-62 [FREE Full text] [doi: 10.1177/1367493518777294] [Medline: 29804471]
74. Baker S, Richards D, Caldwell P. Putting a new intelligent virtual face on a medical treatment advice system to improve adherence. In: Proceedings of the 2014 Conference on Interactive Entertainment. 2014 Presented at: IE2014; December 2-3, 2014; Newcastle, Australia. [doi: 10.1145/2677758.2677762]
75. Richards D, Caldwell P. Building a working alliance with a knowledge based system through an embodied conversational agent. In: Proceedings of the 14th Pacific Rim Knowledge Acquisition Workshop. 2016 Presented at: PKAW 2016; August 22-23, 2016; Phuket, Thailand. [doi: 10.1007/978-3-319-42706-5_16]
76. Richards D, Caldwell P. Improving health outcomes sooner rather than later via an interactive website and virtual specialist. IEEE J Biomed Health Inform 2018 Sep;22(5):1699-1706. [doi: 10.1109/JBHI.2017.2782210] [Medline: 29990229]
77. Klijn AJ, Uiterwaal CS, Vijverberg MA, Winkler PL, Dik P, de Jong TP. Home uroflowmetry biofeedback in behavioral training for dysfunctional voiding in school-age children: a randomized controlled study. J Urol 2006 Jun;175(6):2263-2268. [doi: 10.1016/s0022-5347(06)00331-4]
78. Urbonavičius BG, Kaškonas P. Urodynamic measurement techniques: a review. Measurement 2016 Aug;90:64-73. [doi: 10.1016/j.measurement.2016.04.048]
79. Tokar B, Baskaya M, Celik O, Cemrek F, Acikgoz A. Application of machine learning techniques for enuresis prediction in children. Eur J Pediatr Surg 2021 Oct 20;31(5):414-419. [doi: 10.1055/s-0040-1715655] [Medline: 32820495]

XSL•FO
RenderX

80. Macnab AJ, Shadgan B, Stothers L. The evolution of wireless near infrared spectroscopy applications in urology and rationale for clinical use. J Near Infrared Spectrosc 2012 Jan 01;20(1):57-73. [doi: 10.1255/jnirs.963]
81. Lee DG, Gerber J, Bhatia V, Janzen N, Austin PF, Koh CJ, et al. A prospective comparative study of mobile acoustic uroflowmetry and conventional uroflowmetry. Int Neurourol J 2021 Dec;25(4):355-363 [FREE Full text] [doi: 10.5213/inj.2142154.077] [Medline: 34991305]
82. Schulman SL, Von Zuben FC, Plachter N, Kodman-Jones C. Biofeedback methodology: does it matter how we teach children how to relax the pelvic floor during voiding? J Urol 2001 Dec;166(6):2423-2426. [doi: 10.1016/s0022-5347(05)65607-8] [Medline: 11696803]
83. Sugar EC, Firlit CF. Urodynamic biofeedback: a new therapeutic approach for childhood incontinence/infection (vesical voluntary sphincter dyssynergia). J Urol 1982 Dec;128(6):1253-1258. [doi: 10.1016/s0022-5347(17)53450-3] [Medline: 7154180]
84. Ferroni MC, Chaudhry R, Shen B, Chermansky CJ, Cannon GM, Schneck FX, et al. Transcutaneous electrical nerve stimulation of the foot: results of a novel at-home, noninvasive treatment for nocturnal enuresis in children. Urology 2017 Mar;101:80-84. [doi: 10.1016/j.urology.2016.10.023] [Medline: 27793654]
85. Manson G. A low cost portable enuresis alarm. J Med Eng Technol 1979 Jul 09;3(4):171. [doi: 10.3109/03091907909161670] [Medline: 490604]
86. Parkova I, Valisevskis A, Uzans A, Vilumsone A. Analysis and improvement of nocturnal enuresis alarm system. Industria Textila 2013;64(6):355-361 [FREE Full text]
87. Fordham KE, Meadow SR. Controlled trial of standard pad and bell alarm against mini alarm for nocturnal enuresis. Arch Dis Child 1989 May 01;64(5):651-656 [FREE Full text] [doi: 10.1136/adc.64.5.651] [Medline: 2658853]
88. Peck B, Peck B, Harvey J, Green A, Svedas K, Whitaker S, et al. Body-worn versus bell-and-pad alarm device for the management of monosymptomatic nocturnal enuresis in children: a randomized controlled trial. J Wound Ostomy Continence Nurs 2020;47(5):507-512. [doi: 10.1097/WON.0000000000000678] [Medline: 32970035]
89. Mruzek DW, McAleavey S, Loring WA, Butter E, Smith T, McDonnell E, et al. A pilot investigation of an iOS-based app for toilet training children with autism spectrum disorder. Autism 2019 Feb 07;23(2):359-370. [doi: 10.1177/1362361317741741] [Medline: 29212345]
90. Caldwell PH, Sureshkumar P, Kerr MI, Hamilton S, Teixeira-Pinto A, Macaskill P, et al. A randomised controlled trial of a code-word enuresis alarm. Arch Dis Child 2016 Apr 08;101(4):326-331. [doi: 10.1136/archdischild-2015-308564] [Medline: 26646854]
91. Barroso UJ, Lordêlo P, Teles A, Silveira DM, Renson C, Hoebeke P. New device and new concept for treating nocturnal enuresis: preliminary results of a phase one study. J Pediatr Urol 2014 Dec;10(6):1273-1276. [doi: 10.1016/j.jpurol.2014.06.023] [Medline: 25189504]
92. Elinder G, Soback S. Effect of Uristop on primary nocturnal enuresis. A prospective randomized double-blind study. Acta Paediatr Scand 1985 Jul 21;74(4):574-578. [doi: 10.1111/j.1651-2227.1985.tb11032.x] [Medline: 4040699]
93. Petrican P, Sawan M. Design of a miniaturized ultrasonic bladder volume monitor and subsequent preliminary evaluation on 41 enuretic patients. IEEE Trans Rehabil Eng 1998 Mar;6(1):66-74. [doi: 10.1109/86.662622] [Medline: 9535525]
94. Pretlow RA. Treatment of nocturnal enuresis with an ultrasound bladder volume controlled alarm device. J Urol 1999 Sep;162(3 Part 2):1224-1228. [doi: 10.1016/s0022-5347(01)68141-2]
95. Beauchamp-Parent P, Sawan M. New reconfigurable ultrasonic enuresis monitoring system. In: Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 1998 Presented at: IEMBS 1998; November 1, 1998; Hong Kong, China. [doi: 10.1109/iembs.1998.745549]
96. Kristiansen NK, Djurhuus JC, Nygaard H. Design and evaluation of an ultrasound-based bladder volume monitor. Med Biol Eng Comput 2004 Nov;42(6):762-769. [doi: 10.1007/bf02345209]
97. Padmapriya B, Kesavamurthy T. Implementation of signal processing techniques to measure the urinary bladder volume. J Med Imaging Health Inform 2015 Feb 01;5(1):79-83. [doi: 10.1166/jmihi.2015.1361]
98. Kuru K, Ansell D, Jones M, De Goede C, Leather P. Feasibility study of intelligent autonomous determination of the bladder voiding need to treat bedwetting using ultrasound and smartphone ML techniques: intelligent autonomous treatment of bedwetting. Med Biol Eng Comput 2019 May 26;57(5):1079-1097 [FREE Full text] [doi: 10.1007/s11517-018-1942-9] [Medline: 30588575]
99. Caswell N, Kuru K, Ansell D, Jones MJ, Watkinson BJ, Leather P, et al. Patient engagement in medical device design: refining the essential attributes of a wearable, pre-void, ultrasound alarm for nocturnal enuresis. Pharmaceut Med 2020 Feb 23;34(1):39-48. [doi: 10.1007/s40290-019-00324-w] [Medline: 31970684]
100. Kuru K, Ansell D, Jones M, Watkinson BJ, Caswell N, Leather P, et al. Intelligent autonomous treatment of bedwetting using non-invasive wearable advanced mechatronics systems and MEMS sensors: intelligent autonomous bladder monitoring to treat NE. Med Biol Eng Comput 2020 May 24;58(5):943-965 [FREE Full text] [doi: 10.1007/s11517-019-02091-x] [Medline: 32090271]
101. van Leuteren PJ, Nieuwhof-Leppink AJ, Dik P. SENS-U: clinical evaluation of a full-bladder notification - a pilot study. J Pediatr Urol 2019 Aug;15(4):381.e1-381.e5. [doi: 10.1016/j.jpurol.2019.04.006] [Medline: 31147291]

XSL•FO
RenderX

102. van Leuteren PG, Klijn AJ, de Jong TP, Dik P. SENS-U: validation of a wearable ultrasonic bladder monitor in children during urodynamic studies. J Pediatr Urol 2018 Dec;14(6):569.e1-569.e6. [doi: 10.1016/j.jpurol.2018.07.018] [Medline: 30195717]
103. Kwinten WM, van Leuteren PG, van Duren-van Iersel M, Dik P, Jira PE. SENS-U: continuous home monitoring of natural nocturnal bladder filling in children with nocturnal enuresis - a feasibility study. J Pediatr Urol 2020 Apr;16(2):196.e1-196.e6. [doi: 10.1016/j.jpurol.2020.01.012] [Medline: 32160977]
104. van Leuteren PG, de Vries BA, de Joode-Smink GC, ten Haken B, de Jong TP, Dik P. URIKA, continuous ultrasound monitoring for the detection of a full bladder in children with dysfunctional voiding: a feasibility study. Biomed Phys Eng Express 2017 Feb 06;3(1):017005. [doi: 10.1088/2057-1976/aa589f]
105. Moon J, Lee S, Lee T, Kye S, Lee YS, Shin SC. A prospective study of examining physiological signals for estimating occurrence of nocturnal enuresis. In: Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2017 Presented at: EMBC 2017; July 11-15, 2017; Jeju, Korea. [doi: 10.1109/embc.2017.8037329]
106. Erhel S, Jamet E. Digital game-based learning: impact of instructions and feedback on motivation and learning effectiveness. Comput Educ 2013 Sep;67:156-167. [doi: 10.1016/j.compedu.2013.02.019]
107. Boendermaker WJ, Prins PJ, Wiers RW. Cognitive bias modification for adolescents with substance use problems--can serious games help? J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):13-20 [FREE Full text] [doi: 10.1016/j.jbtep.2015.03.008] [Medline: 25843611]
108. Berg I, Fielding D, Meadow R. Psychiatric disturbance, urgency, and bacteriuria in children with day and night wetting. Arch Dis Child 1977 Aug 01;52(8):651-657 [FREE Full text] [doi: 10.1136/adc.52.8.651] [Medline: 921313]
109. Vasbinder EC, Goossens LM, Rutten-van Mölken MP, de Winter BC, van Dijk L, Vulto AG, et al. e-Monitoring of Asthma Therapy to Improve Compliance in children (e-MATIC): a randomised controlled trial. Eur Respir J 2016 Sep 26;48(3):758-767 [FREE Full text] [doi: 10.1183/13993003.01698-2015] [Medline: 27230437]
110. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. J Consult Clin Psychol 2000;68(3):438-450. [doi: 10.1037//0022-006x.68.3.438]
111. Monkaresi H, Calvo RA, Pardo A, Chow K, Mullan B, Lam M, et al. Intelligent diabetes lifestyle coach. In: Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration. 2013 Presented at: OzCHI '13; November 25-29, 2013; Adelaide, Australia.
112. Yang SS, Wang CC, Chen YT. Home uroflowmetry for the evaluation of boys with urinary incontinence. J Urol 2003 Apr;169(4):1505-1507. [doi: 10.1097/01.ju.0000054761.31086.46] [Medline: 12629404]
113. Nitti VW. Pressure flow urodynamic studies: the gold standard for diagnosing bladder outlet obstruction. Rev Urol 2005;7 Suppl 6(Suppl 6):S14-S21 [FREE Full text] [Medline: 16986024]
114. Walker RM, Di Pasquale B, Hubregtse M, St Clair Carter S. Pressure-flow studies in the diagnosis of bladder outlet obstruction: a study comparing suprapubic and transurethral techniques. Br J Urol 1997 May 29;79(5):693-697. [doi: 10.1046/j.1464-410x.1997.00143.x] [Medline: 9158503]
115. Awad A, Trenfield SJ, Pollard TD, Ong JJ, Elbadawi M, McCoubrey LE, et al. Connected healthcare: improving patient care using digital health technologies. Adv Drug Deliv Rev 2021 Nov;178:113958. [doi: 10.1016/j.addr.2021.113958] [Medline: 34478781]
116. Li J, Carayon P. Health care 4.0: a vision for smart and connected health care. IISE Trans Healthc Syst Eng 2021 Feb 15;11(3):171-180 [FREE Full text] [doi: 10.1080/24725579.2021.1884627] [Medline: 34497970]
117. Gunasekeran DV, Tham YC, Ting DS, Tan GS, Wong TY. Digital health during COVID-19: lessons from operationalising new models of care in ophthalmology. Lancet Digit Health 2021 Feb;3(2):e124-e134. [doi: 10.1016/s2589-7500(20)30287-9]
118. Hassanalieragh M, Page A, Soyata T, Sharma G, Aktas M, Mateos G. Health monitoring and management using internet-of-things (IoT) sensing with cloud-based processing: opportunities and challenges. In: Proceedings of the IEEE International Conference on Services Computing. 2015 Presented at: SCC 2015; June 27-July 2, 2015; New York, NY. [doi: 10.1109/scc.2015.47]
119. Fellah Arbi K, Soulimane O, Saffih F. IoT technologies combining glucose control with physiological signal: comparative study. In: Proceedings of the International Conference on Electrical Engineering. 2020 Presented at: ICEE 2020; September 25-27, 2020; Istanbul, Turkey. [doi: 10.1109/icee49691.2020.9249843]
120. Abdul Minaam DS, Abd-ELfattah M. Smart drugs: improving healthcare using smart pill box for medicine reminder and monitoring system. Future Comput Inform J 2018 Dec;3(2):443-456. [doi: 10.1016/j.fcij.2018.11.008]
121. Kumar SB, Goh WW, Balakrishnan S. Smart medicine reminder device for the elderly. In: Proceedings of the Fourth International Conference on Advances in Computing, Communication & Automation. 2018 Presented at: ICACCA 2018; October 26-28, 2018; Subang Jaya, Malaysia. [doi: 10.1109/icaccaf.2018.8776734]
122. Bladt L, Kashtiara A, Platteau W, De Wachter S, De Win G. First-year experience of managing urology patients with home uroflowmetry: descriptive retrospective analysis. JMIR Form Res 2023 Oct 17;7:e51019 [FREE Full text] [doi: 10.2196/51019] [Medline: 37847531]

XSL•FO
RenderX

123. Bladt L, Blomme RF, Nieuwhof-Leppink AJ, Vermandel A, De Win G, Van Campenhout L. Design of a healthcare ecosystem to improve user experience in pediatric urotherapy. Proc Des Soc 2024 May 16;4:1567-1576. [doi: 10.1017/pds.2024.159]
124. 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) (Text with EEA relevance). European Union. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng [accessed 2025-01-03]
125. Stiglic N, Viner RM. Effects of screentime on the health and well-being of children and adolescents: a systematic review of reviews. BMJ Open 2019 Jan 03;9(1):e023191 [FREE Full text] [doi: 10.1136/bmjopen-2018-023191] [Medline: 30606703]
126. Mendelsohn AL, Brockmeyer CA, Dreyer BP, Fierman AH, Berkule-Silberman SB, Tomopoulos S. Do verbal interactions with infants during electronic media exposure mitigate adverse impacts on their language development as toddlers? Infant Child Dev 2010 Nov 22;19(6):577-593 [FREE Full text] [doi: 10.1002/icd.711] [Medline: 21593996]
127. Christakis DA, Garrison MM, Herrenkohl T, Haggerty K, Rivara FP, Zhou C, et al. Modifying media content for preschool children: a randomized controlled trial. Pediatrics 2013 Mar;131(3):431-438 [FREE Full text] [doi: 10.1542/peds.2012-1493] [Medline: 23420911]
128. Gates M, Hartling L, Shulhan-Kilroy J, MacGregor T, Guitard S, Wingert A, et al. Digital technology distraction for acute pain in children: a meta-analysis. Pediatrics 2020 Feb;145(2):e20191139. [doi: 10.1542/peds.2019-1139] [Medline: 31969473]
129. Linebarger DL, Walker D. Infants' and toddlers' television viewing and language outcomes. Am Behav Sci 2005 Jan 01;48(5):624-645. [doi: 10.1177/0002764204271505]
130. Marsh J, Plowman L, Yamada-Rice D, Bishop JC, Lahmar J, Scott F, et al. Exploring play and creativity in pre-schoolers' use of apps: report for the children's media industry. Technology and Play. 2015. URL: https://www.sheffield.ac.uk/media/28761/download?attachment [accessed 2025-01-03]
131. Vijakkhana N, Wilaisakditipakorn T, Ruedeekhajorn K, Pruksananonda C, Chonchaiya W. Evening media exposure reduces night-time sleep. Acta Paediatr 2015 Mar 29;104(3):306-312. [doi: 10.1111/apa.12904] [Medline: 25521612]
132. Salti R, Tarquini R, Stagi S, Perfetto F, Cornélissen G, Laffi G, et al. Age-dependent association of exposure to television screen with children's urinary melatonin excretion? Neuro Endocrinol Lett 2006;27(1-2):73-80. [Medline: 16648813]
133. Girela-Serrano BM, Spiers AD, Ruotong L, Gangadia S, Toledano MB, Di Simplicio M. Impact of mobile phones and wireless devices use on children and adolescents' mental health: a systematic review. Eur Child Adolesc Psychiatry 2024 Jun 16;33(6):1621-1651 [FREE Full text] [doi: 10.1007/s00787-022-02012-8] [Medline: 35705765]
134. Bladt L, Nieuwhof-Leppink A, Van den Broeck L, Raes T, Wille M, Van Dooren E, et al. The design of an engaging focus group discussion toolkit involving school-aged children following urotherapy. In: Connectivity and Creativity in Times of Conflict. Washington, DC: Academia Press; 2023. [doi: 10.26530/9789401496476-078]

Abbreviations

BBD: bladder and bowel dysfunction
DUI: daytime urinary incontinence
DV: dysfunctional voiding
EMG: electromyography
NE: nocturnal enuresis
OAB: overactive bladder
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial
RPM: remote patient monitoring
UI: urinary incontinence
UTI: urinary tract infection

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 16.09.24; peer-reviewed by E Gea-Izquierdo, K Cakici; comments to author 07.12.24; revised version received 05.03.25; accepted 09.04.25; published 05.05.25.

Please cite as:
Bladt L, Vermeulen J, Vermandel A, De Win G, Van Campenhout L
Innovative, Technology-Driven, Digital Tools for Managing Pediatric Urinary Incontinence: Scoping Review
Interact J Med Res 2025;14:e66336
URL: https://www.i-jmr.org/2025/1/e66336
doi:10.2196/66336
PMID:

©Lola Bladt, Jiri Vermeulen, Alexandra Vermandel, Gunter De Win, Lukas Van Campenhout. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 05.05.2025. 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

Ethical Challenges and Opportunities of AI in End-of-Life Palliative Care: Integrative Review

Abel García Abejas[1,2], MD; David Geraldes Santos[3], PhD; Fabio Leite Costa[1], MD; Aida Cordero Botejara[1], MD; Helder Mota-Filipe[4], PhD; Àngels Salvador Vergés[5], MD, PhD

[1]Faculty of Health Sciences, University of Beira Interior, Lisbon, Portugal

[2]Palliative care, Hospital Lusíadas Lisboa, Lisbon, Portugal

[3]Department of Communication, Philosophy and Politics, University of Beira Interior, Lisbon, Portugal

[4]Pharmacology and Health Technologies, Faculty of Pharmacy, University of Lisbon, Lisbon, Portugal

[5]Board of Directors, Innohealth.Academy, Barcelona, Spain

Corresponding Author:
Àngels Salvador Vergés, MD, PhD
Board of Directors
Innohealth.Academy
Magi Casanovas 11, Sitges
Barcelona, 08870
Spain
Phone: 34 630887995
Email: angels.salvador@innohealth.academy

Abstract

Background: Artificial intelligence (AI) is increasingly integrated into palliative medicine, offering opportunities to improve quality, efficiency, and patient-centeredness in end-of-life care. However, its use raises complex ethical issues, including privacy, equity, dehumanization, and decision-making dilemmas.

Objective: We aim to critically analyze the main ethical implications of AI in end-of-life palliative care and examine the benefits and risks. We propose strategies for ethical and responsible implementation.

Methods: We conducted an integrative review of studies published from 2020 to 2025 in English, Portuguese, and Spanish, identified through systematic searches in PubMed, Scopus, and Google Scholar. Inclusion criteria were studies addressing AI in palliative medicine focusing on ethical implications or patient experience. Two reviewers independently performed study selection and data extraction, resolving discrepancies by consensus. The quality of the papers was assessed using the Critical Appraisal Skills Programme checklist and the Hawker et al tool.

Results: Six key themes emerged: (1) practical applications of AI, (2) communication and AI tools, (3) patient experience and humanization, (4) ethical implications, (5) quality of life perspectives, and (6) challenges and limitations. While AI shows promise for improving efficiency and personalization, consolidated real-world examples of efficiency and equity remain scarce. Key risks include algorithmic bias, cultural insensitivity, and the potential for reduced patient autonomy.

Conclusions: AI can transform palliative care, but implementation must be patient-centered and ethically grounded. Robust policies are needed to ensure equity, privacy, and humanization. Future research should address data diversity, social determinants, and culturally sensitive approaches.

(Interact J Med Res 2025;14:e73517) doi:10.2196/73517

KEYWORDS

palliative medicine; artificial intelligence (AI); ethical implications; end-of-life care; patient-centred care

XSL•FO
RenderX

Introduction

Background

The review focused on studies published between 2020 and 2025 to capture the most recent advances in artificial intelligence (AI) technologies and their application in clinical practice, as the field has evolved rapidly in the last 5 years [1]. This approach ensures the relevance of findings to current and emerging ethical challenges.

AI is a subject within computer science that develops systems capable of performing tasks that simulate human capabilities, such as learning, reasoning, and decision-making. Machine learning (ML) allows algorithms to learn from data without explicit programming within this field. At the same time, deep learning, a subset of ML, uses advanced neural networks to analyze large volumes of information and generate accurate predictions.

AI significantly transforms the traditional health care paradigm toward an evidence-based and patient-centered model. Its application in areas such as the anticipation of complications, the personalization of treatments, and the optimization of resources has proven to be a key catalyst for improving the quality and efficiency of medical care [2].

Palliative medicine has also begun to benefit from the transformative potential of these technologies. This type of care, aimed at patients with advanced or terminal illnesses, seeks to alleviate physical, emotional, and spiritual suffering while seeking to improve quality of life and promote dignity in the final moments [3]. Palliative care encompasses a wide range of conditions, including advanced-stage oncological diseases (metastatic lung, breast, or pancreatic cancer) and nononcological illnesses such as neurodegenerative disorders (amyotrophic lateral sclerosis and late-stage Parkinson), end-stage organ failures (heart, lung, or renal disease), and severe respiratory conditions (chronic obstructive pulmonary disease). These patients, regardless of their specific diagnosis, share everyday needs: symptom relief, emotional support, and dignity preservation as they approach the end of life. The integration of AI in this sensitive context must, therefore, address the heterogeneity of these conditions while upholding ethical principles.

AI in palliative medicine includes tools, such as predictive models, to identify specific needs, wearable devices to monitor symptoms in real time, and virtual assistants that facilitate communication between patients, carers, and professionals. These innovations promise to improve clinical outcomes and enrich the patient experience by offering more personalized approaches [4]. However, its implementation poses significant ethical challenges due to the inherent vulnerability of patients and the complexity of end-of-life decisions. Thus, when we apply AI in palliative care, we must ensure that these tools do not reduce patients to mere actionable data but reinforce their humanity and dignity, honoring their individuality and right to compassionate and ethically informed care. Furthermore, it is crucial to consider how these technologies may affect human dignity and avoid a possible dehumanization of care [5]. In response to these concerns, various institutions have developed ethical guidelines to evaluate and regulate the responsible use of AI-based systems in sensitive contexts.

Despite enthusiasm for AI's transformative potential, significant barriers to its widespread adoption in clinical practice remain. The lack of clear regulatory frameworks and consolidated examples of success highlights the urgent need for integrative research that addresses both the opportunities and the ethical and practical limitations of using AI in palliative care. As Miralles [2] points out in 2023, although multiple promising areas for applying AI in health care have been identified, few consolidated cases have achieved effective adoption in real clinical environments.

In recent years, various ethical self-assessment tools have been developed to verify the suitability of a system based on different ethical principles. In response to these concerns, various institutions have developed ethical guidelines to assess and regulate the responsible use of AI-based systems in sensitive contexts, such as the Ethical Guidelines for Trustworthy AI, the Draft Recommendation on the Ethics of Artificial Intelligence, and [6] the Barcelona Declaration [7]. Also, the AI Ethical Impact Group: From Principles to Practice [8], Technical Methods for Regulatory Inspection of Algorithmic Systems on Social Networking Platforms [9], or the Organisation for Economic Co-Operation and Development (OECD) of AI Systems Classification Framework [10], among others.

In line with international frameworks such as the Institute of Medicine (IOM) and the OECD, this review adopts a multidimensional understanding of "quality of care," which includes safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity as interrelated domains. While "efficiency" is thus an integral component of overall quality, for analytical clarity, we will at times refer to efficiency as system-level performance (resource optimization and process automation) and quality as patient-centered outcomes (dignity and symptom relief). This distinction, while recognizing their overlap, allows us to examine the specific effects of AI on both system operations and patient experience in palliative care.

Theoretical Justification

Ethical reflection on palliative care and AI is rooted in classical and contemporary philosophy.

In his *Nicomachean Ethics* [11], Aristotle posits the notion of the "good life" as the realization of the highest human capacities through virtue, wisdom, and justice. From this perspective, a "good death" implies respecting the dignity and well-being of the patient even at the end of life.

For his part, Immanuel Kant, in the Foundation of the *Metaphysics of Morals* [12], argues that human dignity is an intrinsic and inalienable value, which precludes treating people as mere means to an end, even in medical or technological contexts; this requires that any intervention, including the application of AI, respects the autonomy and inherent value of each patient.

Finally, Emmanuel Lévinas, in *Totality and Infinity* [13], introduces the ethics of otherness, which stresses the importance

XSL•FO
RenderX

of recognizing and preserving the uniqueness of the other. This approach is particularly relevant in palliative care, where care must focus on the individuality and dignity of the patient, avoiding technological reductionism that can depersonalize the end-of-life experience.

These 3 philosophical frameworks provide a sound basis for critically analyzing the opportunities and ethical challenges of integrating AI in palliative care.

We hypothesize that the application of AI in palliative medicine simultaneously offers significant opportunities for personalizing care and presents ethical risks that may compromise patient dignity. This study seeks to explore and examine this hypothesis through an integrative analysis of the recent literature.

Objectives

1. We aim to examine current and potential applications of AI in palliative care: In this systematic review of the literature and recent cases, we will attempt to identify how AI is being used in end-of-life care, including tools for symptom management, clinical decision support, and communication between professionals, patients, and families.
2. To analyze the ethical implications of using AI in palliative care, we will investigate the ethical dilemmas arising from the integration of intelligent technologies in this field, such as privacy and handling of sensitive data, patient autonomy, equity in access to technology, and the possibility of depersonalizing care.
3. We aim to assess the impact of AI on patient experience and dignity at the end of life: To assess how the presence of AI influences the perception of quality of life, respect for dignity, and satisfaction of the emotional and spiritual needs of patients and their families.
4. We aim to propose recommendations for the ethical implementation of AI in palliative care.

Methods

Study Design

This research was carried out as an integrative review, which allows for synthesizing information from various study designs and offers a wide viewpoint on a challenging subject. The review included studies published in Spanish, Portuguese, and English between 2020 and January 2025. Due to their applicability in the biological and technological domains, the scientific databases consulted were PubMed, Scopus, and Google Scholar. Given the rapid development of the field over the past 5 years, we chose to focus on this recent period to ensure the inclusion of the most up-to-date and relevant advancements in AI applied to palliative medicine. This decision is supported by recent systematic reviews documenting a significant increase in the number and scope of published studies in this area [14].

Search and Selection Process

The search strategy was designed to ensure a rigorous and systematic approach. Keywords such as “artificial intelligence,” “palliative care,” “palliative medicine,” “medical ethics,” “machine learning,” and related combinations were used (Multimedia Appendix 1). The search process followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses; Multimedia Appendix 2) guidelines, which provide a standardized framework for transparent and comprehensive reporting, including a 29-item checklist and a 4-phase flow diagram to document the identification, screening, eligibility, and inclusion of studies according to Page et al [15].

Study eligibility was assessed in 2 stages:

- Title and abstract screening: Two reviewers screened all papers and abstracts for relevance.
- Full-text review: The same reviewers independently assessed the full texts of potentially eligible studies.

Two reviewers independently conducted both the study selection and the quality assessment processes. Any discrepancies between reviewers were resolved through consensus discussions or, if necessary, by consulting a third reviewer.

Inclusion criteria were (1) studies published between 2020 and 2025, (2) research addressing the use of AI in palliative medicine, and (3) articles analyzing ethical implications or the patient experience in this context.

Exclusion criteria were (1) studies not explicitly focused on palliative medicine, (2) research lacking ethical analysis or patient experience analysis, and (3) duplicate or non–peer-reviewed publications.

The PRISMA flow diagram summarizes the study selection process, including the number of records identified, screened, excluded, and included at each stage.

Data Extraction

Data extraction was performed independently by 2 reviewers (AGA and ASV). Any discrepancies were discussed and resolved collaboratively. Extracted data included study design, population, AI application, ethical focus, primary findings, and limitations.

Quality Appraisal

The quality of the included studies was assessed using 2 tools to ensure methodological rigor. Two reviewers performed the quality assessment independently, with discrepancies resolved by consensus.

The Critical Appraisal Skills Programme (CASP) checklist [16] was applied as a complementary tool to further assess each study’s methodological quality and transparency (Table 1).

XSL•FO
RenderX

Table 1. Critical appraisal of included studies using CASP[a] checklist.

Study area and author (year)	Key contribution	CASP rating
Prediction and clinical decision-making		
Balch et al (2024) [17]	Review of AI[b] for predicting PRO[c] measures	Medium
Strand et al (2024) [18]	AI/ML[d] model to identify hospitalized patients with cancer needing palliative care	High
He et al (2024) [19]	Effective in targeting palliative support	High
Liu et al (2023) [20]	Accurate prediction of short-term mortality	Medium
Heinzen et al (2023) [21]	Improved early referral to palliative care	High
Morgan et al (2022) [22]	AI improved early identification	High
Porter et al (2020) [23]	Critical reflection on risks or opportunities of AI prediction in palliative care	Medium
Symptom management and quality of life		
Salama et al (2024) [24]	A systematic review of AI/ML in cancer pain management	High
Lazris et al (2024) [25]	Comparison of AI-generated content (ChatGPT) vs NCCN[e] guidelines for cancer symptoms	Medium
Ott et al (2023) [26]	Impact of smart sensors on the “total care“ principle in palliative care	Medium
Deutsch et al (2023) [27]	Improved monitoring and better symptom tracking	High
Yang et al (2021) [28]	Wearables and ML to predict 7-day mortality in terminal cancer	Medium
Communication and emotional support		
Gondode et al (2024) [29]	Performance of AI chatbots (ChatGPT vs Gemini) in palliative care education	Medium
Srivastava and Srivastava (2023) [30]	GPT-3’s potential to improve palliative care communication	Medium
Process automation and modeling		
Reason et al (2024) [31]	LLMs[f] for automating economic modeling in health care	Low
Kamdar et al (2020) [32]	Debate on AI’s future in palliative care or hospice (benefits vs risks)	Medium
Windisch et al (2020) [3]	AI’s role in improving the timing or quality of palliative interventions	Medium
Ethics and challenges		
See (2024) [33]	AI as an ethical advisor in clinical contexts	Medium
Adegbesan et al (2024) [34]	Ethical challenges of AI integration in palliative care	High
Ranard et al (2024) [35]	Minimizing algorithmic biases in critical care via AI	High
De Panfilis et al (2023) [36]	Framework for future policy	High
Ferrario et al (2023) [5]	Ethics of algorithmic prediction of end-of-life preferences	High
Meier et al (2022) [37]	Framework for ethical algorithm-based clinical decisions	Medium
Research and review of advances		
Bozkurt et al (2024) [1]	Protocol for assessing AI data diversity in palliative care	Medium
Macheka et al (2024) [38]	Prospective assessment of AI in postdiagnostic cancer care. High feasibility and good patient feedback.	High
Vu et al (2023) [14]	Systematic review of ML applications in palliative care	High
Reddy et al (2023) [39]	Review of AI advances for palliative cancer care	Medium
Barry et al (2023) [40]	Challenges for evidence-based palliative care delivery	Medium
Chua et al (2021) [41]	Path to AI implementation in oncology; pragmatic roadmap created	Medium

[a]CASP: Critical Appraisal Skills Programme.

[b]AI: artificial intelligence.

[c]PRO: patient-reported outcome.

[d]ML: machine learning.

[e]NCCN: National Comprehensive Cancer Network.

XSL•FO
RenderX

[f]LLM: large language model.

The Hawker et al (2002) [42] checklist allows for systematic evaluation across diverse research designs. Each study was scored independently across 11 domains: clarity of purpose, study design, methodology, sampling, data analysis, ethical implications, relevance, transferability, results, discussion, and theoretical basis. Scores range from 1 (very poor) to 4 (good) for each domain; an overall quality score was calculated as the mean of all domain scores (Table 2).

Table 2. Methodological quality assessment of all included studies according to the criteria of Hawker et al [42]: 1=very poor, 2=poor, 3=fair, and 4=good.

Study	Clarity of purpose	Design	Methodology	Sampling	Analysis	Ethical implications	Relevance	Transferability	Results	Discussion	Theoretical basis	Overall quality
Balch et al [17]	4	4	4	2	3	3	3	4	4	3	3	3.27
Strand et al [18]	4	3	3	2	3	3	3	4	4	3	3	3.09
He et al [19]	2	3	3	2	3	3	3	4	4	3	3	2.91
Liu et al [20]	4	3	3	2	3	3	3	4	4	3	3	3.09
Heinzen et al [21]	4	3	3	2	3	3	3	4	4	3	3	3.09
Morgan et al [22]	4	3	2	3	3	3	3	4	4	3	3	3.09
Porter et al [23]	2	3	3	2	3	3	3	4	4	3	3	2.91
Salama et al [24]	4	3	3	2	3	3	3	4	4	3	3	3.09
Lazris et al [25]	4	3	3	2	3	3	3	4	4	3	3	3.09
Ott et al [26]	4	3	3	2	3	3	3	4	4	3	3	3.09
Deutsch et al [27]	2	3	3	2	3	3	3	4	4	3	3	2.91
Yang et al [28]	4	3	3	2	3	3	3	4	4	3	3	3.09
Gondode et al [29]	4	3	3	2	3	3	3	4	4	3	3	3.09
Srivastava and Srivastava [30]	4	3	3	2	3	3	3	4	4	3	3	3.09
Reason et al [31]	4	3	3	2	3	3	3	4	4	3	3	3.09
Kamdar et al [32]	2	3	3	2	3	3	3	4	4	3	3	2.91
Windisch et al [3]	4	3	3	2	3	3	3	4	4	3	3	3.09
See [33]	4	3	3	2	3	3	3	4	4	3	3	3.09
Adegbesan et al [34]	4	3	3	2	3	3	3	4	4	3	3	3.09
Ranard et al [35]	2	3	3	2	3	3	3	4	4	3	3	2.91
De Panfilis et al [36]	4	3	3	2	3	3	3	4	4	3	3	3.09
Ferrario et al [5]	4	3	3	2	3	3	3	4	4	3	3	3.09
Meier et al [37]	4	3	3	2	3	3	3	4	4	3	3	3.09
Bozkurt et al [1]	2	3	3	2	3	3	3	4	4	3	3	2.91
Macheka et al [38]	4	3	3	2	3	3	3	4	4	3	3	3.09
Vu et al [14]	4	3	3	2	3	3	3	4	4	3	3	3.09
Reddy et al [39]	4	3	3	2	3	3	3	4	4	3	3	3.09
Barry et al [40]	2	3	3	2	3	3	3	4	4	3	3	2.91
Chua et al [41]	4	3	3	2	3	3	3	4	4	3	3	3.09

Thematic Analysis

Extracted data were thematically coded and grouped into 6 key categories—prediction and clinical decision-making, symptom management and quality of life, communication and emotional support, process automation and modeling, ethical implications, and research and review of advances in AI [14]—reflecting the main areas presented in the Results section.

XSL•FO
RenderX

Ensuring Methodological Rigor

To maximize the reliability and validity of the review:

- Triangulation: findings were compared across studies to identify consistent patterns.
- Peer review: the methodology was reviewed using bioethics and AI.
- Critical evaluation: each study was assessed for quality, relevance, and validity using the criteria above.

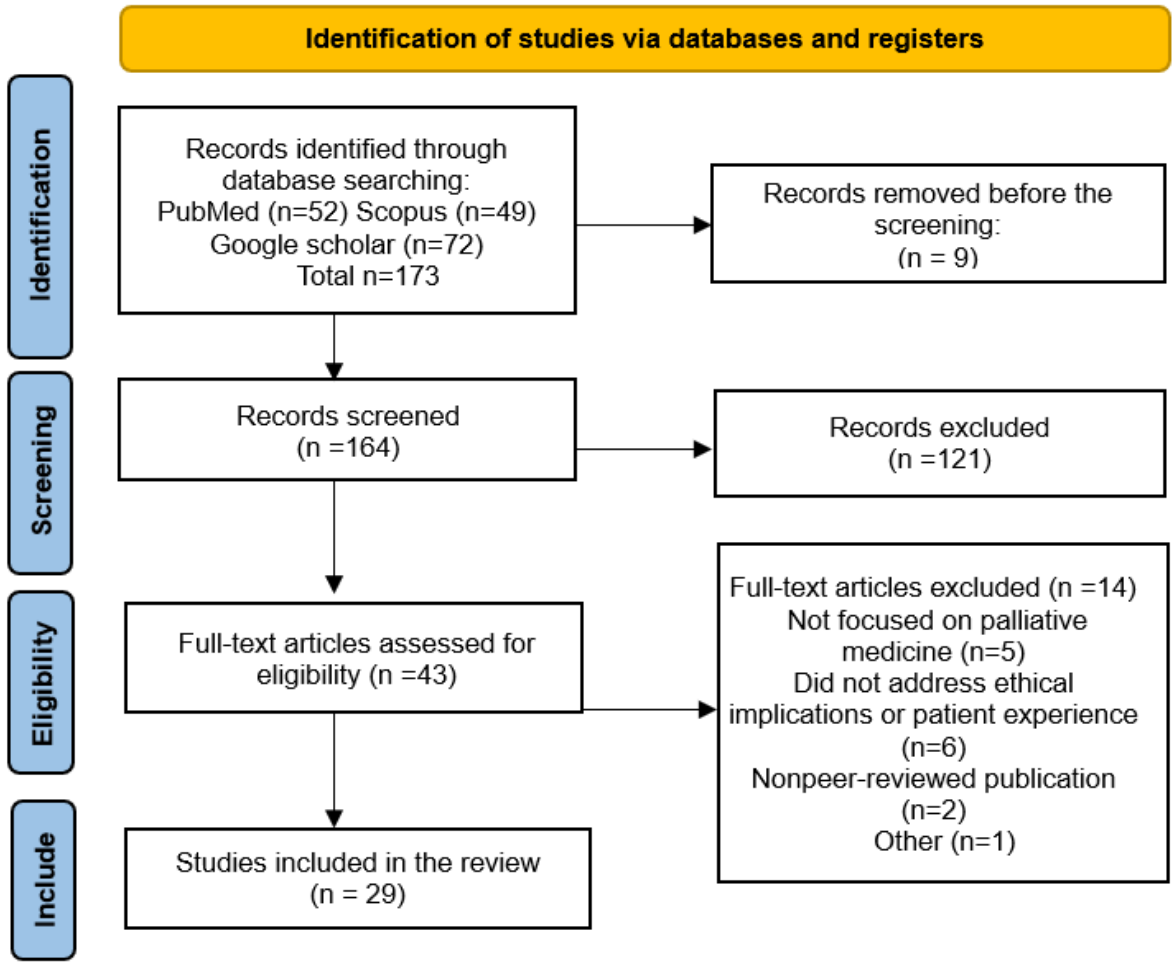
This process allowed identifying strengths, limitations, and potential biases in the included studies [43].

Results

Overview

This section presents a thematic synthesis of the main findings regarding the ethical and practical implications of AI in palliative medicine at the end of life. The 29 included studies, published between 2020 and January 2025, covered various clinical contexts, populations, and AI applications. The results are structured in 6 key areas identified in the literature (Figure 1).

Figure 1. PRISMA 2020 flow diagram illustrating the study selection process. The diagram shows the number of records identified, duplicates removed, records screened, full-text papers assessed, and studies included in the review. Adapted from the PRISMA 2020 Statement [15].



The following tables provide a comprehensive synthesis of the 29 studies included in this review.

In Table 1, we present the ratings obtained for the methodological quality of each study according to the CASP checklist. Two independent reviewers rated each item as "yes," "no," or "unclear," and the percentage of responses was calculated to assign an overall rating of high, medium, or low.

Table 2 shows the quality scores in 9 domains using the Hawker et al [42] instrument. Two reviewers rated each domain from 1 (very poor) to 4 (excellent), and we report both the individual domain scores and the overall mean quality score for each item.

Table 3 summarizes the key characteristics of the included studies. For each study, we list the author, year, country, design, AI/ML application, population or setting, study objective, and principal findings, and group them according to the 5 thematic areas identified in our review.

The analysis of the included studies revealed 6 key thematic areas in the application of AI in palliative care. These thematic areas provide a comprehensive overview of the current landscape and highlight both the opportunities and challenges presented by AI in this field.

XSL•FO
RenderX

Table 3. Characteristics of included studies. The main characteristics of the included studies are by thematic area, design, and key findings.

Author (year)	Country	Study design	AI[a] application	Population or context	Study aim	Key findings
Balch et al (2024) [17]	United States	Review	AI for predicting PROs[b]	Patients with advanced cancer	Explore the use of AI in predicting PROs	AI shows potential but lacks validation
Strand et al (2024) [18]	United States	ML[c] model development	Mortality prediction tool	Hospitalized patients with cancer	Develop a model to identify palliative needs	High predictive value for end-of-life care
He et al (2024) [19]	United States	Cohort study	ML for palliative consultation allocation	Patients with cancer	Assign consultations based on predicted need	Effective in targeting palliative support
Liu et al (2023) [20]	Taiwan	Observational	Wearables and ML	Patients with terminal cancer	Predict mortality risk in real time	Accurate prediction of short-term mortality
Heinzen et al (2023) [21]	Germany	RCT[d]	ML timing intervention	Primary care	Assess the impact on care timing	Improved early referral to palliative care
Morgan et al (2022) [22]	United States	RCT	AI prediction of care needs	Advanced cancer	Evaluate AI vs traditional triage	AI improved early identification
Porter et al (2020) [23]	United Kingdom	Critical reflection	Ethical analysis of prediction	General palliative care	Reflect on risks and values in AI prediction	Warns about the dehumanization risk
Salama et al (2024) [24]	United States	Systematic review	ML in pain management	Patients with cancer	Evaluate AI effectiveness in pain treatment	Supports the integration of AI tools
Lazris et al (2024) [25]	United States	Comparison study	ChatGPT vs NCCN[e]	Cancer symptom guidance	Evaluate content quality	AI is aligned with guidelines in most areas
Ott et al (2023) [21]	Germany	Observational	Smart sensors for monitoring	Palliative care patients	Assess the "total care" principle using technology	Improved monitoring and better symptom tracking
Deutsch et al (2023) [27]	Germany	Observational	ML for PROs monitoring	Metastatic cancer	Track patient outcomes	Improved reporting and early alerts
Yang et al (2021) [28]	China	Cohort study	Wearables and ML	Terminal cancer	Predict 7-day mortality	High predictive accuracy
Gondode et al (2024) [29]	United States	Comparative study	ChatGPT vs Gemini in education	Health care professionals	Compare chatbot effectiveness	Both tools are effective; Gemini is more accurate
Srivastava and Srivastava (2023) [30]	India	Exploratory	GPT-3 communication support	General palliative population	Examine AI in patient-clinician dialogue	Potential for improving conversations
Reason et al (2024) [31]	United Kingdom	Implementation study	LLMs[f] for economic modeling	Palliative systems planning	Automate health economic models	LLMs reduce the workload but need oversight
Kamdar et al (2020) [32]	United States	Debate or commentary	AI in palliative care	General palliative systems	Debate AI's pros and cons	Highlights opportunities and ethical risks
Windisch et al (2020) [3]	Germany	Case study	AI-enhanced timing	Hospital-based care	Improve intervention timing	Faster, more targeted responses
See (2024) [33]	United States	Qualitative	AI as an ethical advisor	Oncology settings	Evaluate AI-generated ethical suggestions	Useful but lacking nuance
Adegbesan et al (2024) [34]	Nigeria	Thematic analysis	Ethical challenges of AI	Low-resource settings	Explore equity and justice issues	AI raises equity concerns
Ranard et al (2024) [35]	United States	Technical study	Bias minimization algorithms	Critical care AI	Reduce bias in predictions	The algorithm reduced disparities in results
De Panfilis et al (2023) [36]	Italy	Conceptual framework	Ethical issues in AI	Palliative care context	Define ethical concerns	Framework for future policy

XSL•FO
RenderX

Author (year)	Country	Study design	AI[a] application	Population or context	Study aim	Key findings
Ferrario et al (2023) [5]	United Kingdom	Ethical analysis	Predictive systems for end-of-life	Hospice settings	Assess algorithmic risks	Need for transparent systems
Meier et al (2022) [37]	United States	Framework proposal	AI clinical decision support	Advanced illness patients	Propose ethical guidance	Applicable for clinical protocol design
Bozkurt et al (2024) [1]	Turkey	Protocol	Diversity metrics in data	Mixed cancer cohorts	Establish a framework for diversity	Supports inclusive data use
Macheka et al (2024) [38]	Zimbabwe	Prospective evaluation	AI in postdiagnosis care	Rural patients with cancer	Evaluate implementation outcomes	High feasibility and good patient feedback
Vu et al (2023) [14]	Switzerland	Systematic review	ML in palliative care	Various populations	Map applications and outcomes	AI is growing in scope and evidence
Reddy et al (2023) [39]	India	Narrative review	AI for palliative oncology	Patients with cancer	Summarize recent AI use	Progress is seen, but fragmented evidence
Barry et al (2023) [40]	United States	Survey	Evidence-based palliative AI	Clinicians and patients	Identify barriers to adoption	Concerns about data quality and trust
Chua et al (2021) [41]	Singapore	Implementation framework	AI in oncology	Urban hospitals	Design path for AI adoption	Pragmatic roadmap created

[a]AI: artificial intelligence.

[b]PRO: patient-reported outcome.

[c]ML: machine learning.

[d]RCT: randomized controlled trial.

[e]NCCN: National Comprehensive Cancer Network.

[f]LLM: large language model.

Prediction and Clinical Decision-Making

AI has demonstrated significant potential in supporting clinical decision-making and anticipating patient needs in palliative care. For example, Strand et al [18] developed a machine learning model that more accurately identified hospitalized patients with cancer who could benefit from specialized palliative care, outperforming traditional approaches. Similarly, Salama et al [24] and Liu et al [20] reported that AI tools can help personalize pain management and predict imminent terminal events, allowing for more timely interventions. However, Porter et al [23] cautioned that excessive reliance on algorithmic predictions, especially when models lack transparency or interpretability, may undermine human sensitivity and clinical judgment in complex palliative care scenarios.

Symptom Management and Quality of Life

Effective symptom management and quality of life improvement are central in palliative medicine, and AI tools are being tested in oncological and nononcological populations. AI applications have facilitated individualized pain control and symptom monitoring for patients with cancer [24]. In noncancer contexts, Ott et al [26] described using smart sensors for real-time symptom tracking in neurodegenerative diseases. However, studies such as by Deutsch et al [27] highlighted the risk of bias when training datasets underrepresent specific populations (patients without cancer or minority groups), potentially limiting the generalizability and fairness of AI-driven symptom management.

Communication and AI Tools

AI-based communication tools, including chatbots and natural language processing models, have been explored to support interactions between professionals, patients, and families. Gondode et al [29] and Srivastava and Srivastava [30] analyzed large language models GPT-3 to facilitate information delivery and emotional support. However, these tools often reflect Western bioethical principles. They may not adapt well to cultural contexts where family-centered decision-making or gradual truth disclosure is preferred [34]. Several studies reported that AI trained on Anglo-Saxon datasets may misinterpret emotional cues or cultural preferences, underlining the need for culturally sensitive models and community co-design.

Process Automation and Modeling

AI process automation can optimize resource allocation and improve efficiency in palliative care. Reason et al [31] demonstrated how large language models can automate economic modeling, potentially reducing costs and improving service access. Windisch et al [3] emphasized the benefits of AI in improving the timing and quality of palliative interventions. However, Kamdar et al [32] stressed the importance of maintaining a patient-centered approach, even in highly automated environments.

Ethical Implications

The ethical challenges of AI in palliative medicine are complex and multifaceted. Ferrario et al [5] analyzed the need for transparency and accountability in algorithmic prediction of end-of-life preferences. Ranard et al [35] addressed the risks of algorithmic bias, especially when models are trained on

XSL•FO
RenderX

unrepresentative data, which can lead to inequitable care decisions. Finally, See [33] explored the potential of AI as an ethical advisor but noted that this application is still in its early stages and requires further research.

Regarding the ethical design of data-driven decision support tools, Bak et al [44] discuss the importance of considering ethical principles, such as algorithmic fairness and privacy, in the development of decision support tools in oncology, with direct implications for palliative medicine.

Furthermore, regarding the ethical design of data-driven decision support tools, Balch et al [17] discuss the importance of considering ethical principles, such as algorithmic fairness and privacy, in developing decision support tools in oncology, with direct implications for palliative medicine.

Ethical considerations should also be considered in cancer chatbots. In their study, Chow et al [45-52] address the need for transparency and informed consent in the use of AI-based chatbots in cancer care, emphasizing the risks of dehumanization and loss of trust.

Research and Review of Advances in AI

Recent reviews and methodological studies have documented advances and limitations in AI applications for palliative care. Vu et al [14] systematically reviewed ML applications, highlighting the need for more robust, real-world evidence. Reddy et al [39] summarized advances in AI-based symptom management, and Bozkurt et al [1] developed protocols to assess the robustness of these applications. Macheka et al [38] evaluated the included role of AI in postdiagnostic treatment pathways, emphasizing the need for continuous research and ethical oversight.

AI offers innovative prediction, symptom management, communication, and resource optimization solutions in palliative medicine. However, its implementation is accompanied by significant ethical, cultural, and practical challenges, especially regarding equity, humanization, and respect for patient autonomy. The literature highlights the importance of addressing patient heterogeneity, cultural context, and social determinants to ensure that AI applications are practical and ethically acceptable.

Discussion

Principal Findings

This integrative review demonstrates that AI is increasingly present in palliative care, offering innovative solutions for clinical prediction, symptom management, communication, and process automation. The main findings suggest that while AI can improve efficiency and support decision-making, there remains a significant lack of consolidated, real-world examples that simultaneously demonstrate both efficiency and equity in outcomes. Most published studies focus on technical feasibility or operational improvements, but few document how AI enhances equitable access or patient-centered outcomes across diverse populations. This gap underscores the need for more robust, contextually grounded evidence to guide the ethical implementation of AI in end-of-life care [36].

A key finding is the persistent tension between efficiency and quality. Although frameworks such as the IOM and OECD recognize efficiency, equity, and patient-centeredness as embedded dimensions of quality, our review shows that improvements in system-level efficiency (resource optimization, automated symptom tracking) do not always translate into perceived improvements in care quality by patients and families. In palliative care, relational and dignity-centered outcomes, such as humanization, emotional support, and respect for autonomy, remain fundamental and may be at risk if AI is implemented without careful ethical consideration [36].

It should be noted that the paper distinguishes between “quality” of care and “efficiency.” In contrast, efficiency, equity, and patient-centeredness are internationally recognized as embedded measures of quality of care. According to frameworks such as the IOM [49], quality of care encompasses 6 interrelated domains: safety, effectiveness, patient-centeredness, timeliness, efficiency, and equity. Thus, efficiency is not a separate attribute but an integral component of quality, alongside equity and patient-centeredness. We recommend harmonizing the terminology and analysis to reflect this international consensus, avoiding an artificial dichotomy between quality and efficiency.

Comparison to Prior Work

Our findings align with previous reviews, highlighting both AI’s transformative potential in palliative settings and the ethical challenges it introduces. Recent literature confirms that AI’s integration in palliative care is still early, with limited robust evidence for improved equity or patient experience. While some studies report promising advances in prediction and symptom management, others caution about algorithmic bias, lack of transparency, and the risk of dehumanization [50].

The review expands on prior work by addressing ethical principles’ historical and cultural variability. Many AI tools in palliative care are developed and validated in Western contexts, reflecting assumptions about autonomy, truth-telling, and individual decision-making that may not be universally applicable. Studies show that AI models trained on Anglo-Saxon datasets can misinterpret emotional cues or cultural preferences, particularly in Southern European, Latin American, or other non-Western settings where family-centered decision-making and gradual truth disclosure are everyday occurrences [34]. This highlights the need for culturally sensitive AI models and participatory design processes.

Real Case: Mortality Prediction and Advance Care Planning

A recent study analyzed the implementation of an AI system designed to predict the likelihood of a patient dying in the next 12 months to facilitate timely discussions about palliative care. The study developed an explainable ML model using electronic health records data to proactively identify patients with advanced cancer at high risk of mortality. The model demonstrated strong predictive performance (area under receiver operating characteristic curve 0.861) and was intended to support early integration of palliative care in outpatient oncology settings [51]. However, introducing these predictions into the clinical process generated significant disagreements among health care

XSL•FO
RenderX

professionals, patients, and family members. The main concerns included:

- Patient autonomy: algorithmic predictions were sometimes used without adequate informed consent, potentially undermining patients' ability to decide about their care.
- Justice and equity: models trained on unrepresentative data risked introducing bias, disproportionately affecting marginalized groups.
- Beneficence and nonmaleficence: overreliance on AI predictions led to inappropriate interventions, such as premature end-of-life planning, without considering individual complexity.

This case highlights the need for a rigorous ethical approach to integrating AI into palliative care, ensuring that technologies complement clinical judgement and human empathy, not replace them.

Strengths and Limitations

This integrative review has several strengths. First, it offers a comprehensive and up-to-date synthesis of the recent literature on AI in end-of-life palliative care, focusing on the period between 2020 and 2025. This time frame was chosen to reflect the most current technological developments and their ethical implications. Second, the review's methodological rigor is reinforced using 2 quality appraisal tools, CASP and Hawker et al [42], which allowed a systematic evaluation across diverse study designs. Additionally, including philosophical, clinical, and ethical perspectives provides a multidimensional framework for understanding the challenges and opportunities of AI in this sensitive field. Furthermore, the review is enriched by presenting a real-world case, illustrating the practical dilemmas and tensions encountered when implementing predictive AI tools in clinical settings.

Nonetheless, several limitations must be acknowledged. The narrowed time frame may have excluded relevant earlier studies; however, this was a deliberate choice to focus on recent developments most applicable to current practice. The language scope was restricted to English, Portuguese, and Spanish, which may limit the generalizability of findings to other cultural contexts. Additionally, the heterogeneity of study designs and the variability in reporting standards made it difficult to perform a meta-analysis, which was replaced by thematic synthesis. Another limitation is that while some studies addressed social determinants of health (SDOH), such as literacy, socioeconomic status, or geographic inequalities, this aspect was not systematically captured across the entire corpus. Finally, given the fast-evolving nature of AI technologies, the findings and conclusions may become outdated over time, highlighting the need for ongoing surveillance and review.

The review of Ortiz et al [53] addressed the opportunities and challenges in using wearable sensors to monitor patients with cancer, including data integration and user acceptance issues.

Future Directions

Integrating AI into palliative care necessitates a balanced approach that prioritizes ethical rigor, patient-centered outcomes, and cultural adaptability. Based on recent evidence, the following directions are critical for advancing this field responsibly.

Enhancing Predictive Models With Diverse Data

ML models, such as those predicting mortality to facilitate early palliative care referrals in Medicare Advantage populations, demonstrate high accuracy but require broader validation across diverse demographics. Future research must prioritize datasets that include underrepresented groups (patients without cancer and ethnic minorities) to mitigate algorithmic bias and ensure equitable access to palliative services. For instance, models trained on SDOH and multi-institutional data could improve generalizability while addressing systemic disparities in end-of-life care [44].

Ethical Co-Design of Decision Support Tools

The development of AI-driven decision support systems in oncology, as exemplified by the 4D PICTURE project, highlights the importance of participatory design involving patients, clinicians, and ethicists [54]. Key strategies include the following: (1) ethical review processes will need to be integrated to address data bias, privacy, and transparency; (2) ensuring AI tools align with cultural norms (family-centered decision-making in non-Western contexts); and (3) adopting frameworks such as design justice to empower marginalized voices in tool development.

Hybrid Care Models Balancing Telehealth and Human Interaction

While telehealth improves accessibility and psychological comfort for palliative patients, studies emphasize that optimal care requires complementing AI tools with in-person interactions [54]. Future implementations should (1) integrate AI for routine monitoring (symptom tracking via wearables) while reserving complex decisions for clinician-patient dialogues and (2) address technical barriers (poor internet connectivity) that undermine interpersonal connection in telehealth.

Transparency and Explainability in Mortality Prediction

Explainable AI models, such as the transparent mortality prediction tool for patients with cancer developed by Bertsimas et al [55], enhance clinician trust and facilitate shared decision-making. Recommendations include (1) standardizing reporting of model features (weight changes and albumin levels) to improve clinical interpretability and (2) validating tools in real-world settings to assess their impact on goals-of-care discussions and patient autonomy.

AI as a Supplement, Not Replacement, for Human Judgment

Despite their reliability in answering breast cancer surgery queries (average score 3.98/5), AI chatbots such as ChatGPT must be framed as supplements to, not substitutes for, medical expertise [56]. Future work should (1) develop guardrails to prevent overreliance on AI, such as mandatory disclaimers urging users to consult clinicians, and (2) train chatbots to recognize cultural nuances in communication (gradual truth disclosure in specific populations).

XSL•FO
RenderX

Ethical Deployment of Oncology Chatbots

Building on frameworks for human-centered AI in oncology [45], developers must prioritize (1) transparency: disclosing data sources and limitations to users, (2) autonomy: allowing patients to opt out of AI-driven interactions, and (3) equity: ensuring chatbots are accessible across literacy and socioeconomic levels.

Conclusions

AI is poised to transform end-of-life palliative care, offering innovative clinical prediction, symptom management, communication, and resource optimization solutions. However, this review highlights that AI's successful and ethical integration in such a sensitive field is far from straightforward.

First, while AI can improve efficiency and operational aspects of care, there is still a notable lack of consolidated, real-world examples demonstrating both efficiency and equity in outcomes. Most current evidence is derived from pilot studies or technical feasibility reports, with limited documentation of long-term, patient-centered benefits across diverse populations.

Second, the ethical challenges surrounding AI in palliative care are complex and multifaceted. Issues such as algorithmic bias, lack of transparency, risks to patient autonomy, and the potential for dehumanization must be addressed proactively. The review underscores that efficiency, equity, and patient-centeredness are not isolated goals but interrelated dimensions of quality care, as recognized by international frameworks such as the IOM.

Third, cultural context and patient heterogeneity are critical factors. Many AI tools are developed based on Western bioethical assumptions, which may not be universally applicable, especially in settings where family-centered decision-making or culturally nuanced understandings of pain and suffering prevail. Addressing SDOH and ensuring the inclusion of marginalized voices in AI development is essential for equitable implementation.

Fourth, patients' and families' experiences must remain at the heart of palliative care innovation. AI should complement, not replace, clinical judgment, human empathy, and the relational aspects that define quality end-of-life care.

Key Recommendations

Our key recommendations are to (1) develop clear policies and regulatory frameworks to ensure fairness, privacy, transparency, and accountability in using AI in palliative care; (2) promote the humanization of care, design AI tools that support, rather than supplant, compassionate human interaction and shared decision-making; (3) foster continuous, multidisciplinary research to rigorously evaluate the benefits and risks of AI, address algorithmic biases, and adapt tools to diverse clinical and cultural contexts; and (4) encourage participatory approaches involving patients, families, clinicians, ethicists, and community representatives in designing, implementing, and evaluating AI systems.

Final Perspective

Ultimately, the ethical adoption of AI in palliative medicine requires a careful balance between technological innovation and the preservation of human dignity. Only through patient-centered, culturally sensitive, and ethically grounded strategies can we maximize the benefits of AI while mitigating its risks, thus improving the experience and outcomes for patients and families at the end of life.

Acknowledgments

The authors would like to thank the individuals whose support facilitated the conduct of this study.

Authors' Contributions

Each author actively participated in this study's conception, development, and writing. AGA, DGS, and ASV contributed to conceptualizing this study and its methodology, while FLC, ACB, and HM-F contributed technical expertise and interpretation of the findings. AGA and ASV participated in drafting this paper and critically revising the content. All authors have read and approved the final version of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Review strategy and quality assessment.
[DOCX File , 18 KB - ijmr_v14i1e73517_app1.docx]

Multimedia Appendix 2
PRISMA checklist.
[DOCX File , 271 KB - ijmr_v14i1e73517_app2.docx]

References

XSL•FO
RenderX

1. Bozkurt S, Fereydooni S, Kar I, Diop Chalmers C, Leslie SL, Pathak R, et al. Investigating data diversity and model robustness of AI applications in palliative care and hospice: protocol for scoping review. JMIR Res Protoc 2024;13:e56353 [FREE Full text] [doi: 10.2196/56353] [Medline: 39378420]
2. Miralles F. La transformación de la salud mediante uso de datos e inteligencia artificial: ámbitos de aplicación, retos científico-tecnológicos y claves para la adopción. Fundació Víctor Grífols i Lucas. 2023. URL: https://www.fundaciogrifols.org/documents/4438882/5272129/Q63_inteligencia_artificial.pdf [accessed 2025-05-03]
3. Windisch P, Hertler C, Blum D, Zwahlen D, Förster R. Leveraging advances in artificial intelligence to improve the quality and timing of palliative care. Cancers (Basel) 2020;12(5):1149 [FREE Full text] [doi: 10.3390/cancers12051149] [Medline: 32375249]
4. Tei S, Fujino J. Artificial intelligence, internet addiction, and palliative care. Eur. Psychiatr 2024;67(S1):S339 [FREE Full text] [doi: 10.1192/j.eurpsy.2024.700]
5. Ferrario A, Gloeckler S, Biller-Andorno N. Ethics of the algorithmic prediction of goal of care preferences: from theory to practice. J Med Ethics 2023;49(3):165-174 [FREE Full text] [doi: 10.1136/jme-2022-108371] [Medline: 36347603]
6. Preliminary report on the first draft of the recommendation on the ethics of artificial intelligence [Internet]. UNESCO. 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000374266 [accessed 2025-05-03]
7. Barcelona Declaration for the proper development and usage of artificial intelligence in Europe [Internet]. IIIA CSIC. 2017. URL: https://www.iiia.csic.es/barcelonadeclaration/ [accessed 2025-05-03]
8. From principles to practice: an interdisciplinary framework to operationalise AI ethics [Internet]. AI Ethics Impact Group. 2020. URL: https://www.ai-ethics-impact.org/en [accessed 2025-05-03]
9. Technical methods for regulatory inspection of algorithmic systems in social media platforms [Internet]. Ada Lovelace Institute. 2021. URL: https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/ [accessed 2025-05-03]
10. OECD. OECD framework for the classification of AI systems. OECD Digit Econ Paper. Organisation for Economic Co-Operation and Development. 2023. URL: https://www.oecd.org/en/publications/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en.html [accessed 2025-05-03]
11. Aristotle. In: Crisp R, editor. Aristotle: Nicomachean Ethics. Cambridge: Cambridge University Press; 2000.
12. Kant I. In: Gregor M, Timmermann J, editors. Immanuel Kant: Groundwork of the Metaphysics of Morals: A German English. Cambridge: Cambridge University Press; 2011.
13. Idareta Goldaracena F, Úriz Peman MJ. Aportaciones de la ética de la alteridad de E. Lévinas y la ética del cuidado de C. Gilligan a la intervención en trabajo social. Alternativas 2012(19):33-44. [doi: 10.14198/ALTERN2012.19.02]
14. Vu E, Steinmann N, Schröder C, Förster R, Aebersold DM, Eychmüller S, et al. Applications of machine learning in palliative care: a systematic review. Cancers (Basel) 2023;15(5):1596. [doi: 10.3390/cancers15051596] [Medline: 36900387]
15. Page MJ, Moher D, McKenzie JE. Introduction to preferred reporting items for systematic reviews and meta-analyses 2020 and implications for research synthesis methodologists. Res Synth Methods 2022;13(2):156-163. [doi: 10.1002/jrsm.1535] [Medline: 34693644]
16. CASP checklists. [Internet]. Critical Appraisal Skills Programme: CASP. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2025-05-03]
17. Balch JA, Chatham AH, Hong PKW, Manganiello L, Baskaran N, Bihorac A, et al. Predicting patient reported outcome measures: a scoping review for the artificial intelligence-guided patient preference predictor. Front Artif Intell 2024;7:1477447 [FREE Full text] [doi: 10.3389/frai.2024.1477447] [Medline: 39564457]
18. Strand JJ, Morgan AA, Pachman DR, Storlie CB, Wilson PM. Performance of an artificial intelligence/machine learning model designed to identify hospitalized patients with cancer who could benefit from timely specialized palliative care delivery. JCO 2024;42(16_suppl):1559-1559 [FREE Full text] [doi: 10.1200/jco.2024.42.16_suppl.1559]
19. He JC, Moffat GT, Podolsky S, Khan F, Liu N, Taback N, et al. Machine learning to allocate palliative care consultations during cancer treatment. J Clin Oncol 2024;42(14):1625-1634. [doi: 10.1200/JCO.23.01291] [Medline: 38359380]
20. Liu JH, Shih CY, Huang HL, Peng JK, Cheng SY, Tsai JS, et al. Evaluating the potential of machine learning and wearable devices in end-of-life care in predicting 7-day death events among patients with terminal cancer: cohort study. J Med Internet Res 2023;25:e47366 [FREE Full text] [doi: 10.2196/47366] [Medline: 37594793]
21. Heinzen EP, Wilson PM, Storlie CB, Demuth GO, Asai SW, Schaeferle GM, et al. Impact of a machine learning algorithm on time to palliative care in a primary care population: protocol for a stepped-wedge pragmatic randomized trial. BMC Palliat Care 2023;22(1):9 [FREE Full text] [doi: 10.1186/s12904-022-01113-0] [Medline: 36737744]
22. Morgan A, Karow J, Olson E, Strand J, Wilson P, Soleimani J, et al. Randomized trial of a novel artificial intelligence/machine learning model to predict the need for specialty palliative care. J Pain Symptom Manage 2022;63(5):879-880 [FREE Full text] [doi: 10.1016/j.jpainsymman.2022.02.078]
23. Porter AS, Harman S, Lakin JR. Power and perils of prediction in palliative care. Lancet 2020;395(10225):680-681. [doi: 10.1016/S0140-6736(20)30318-4] [Medline: 32113499]
24. Salama V, Godinich B, Geng Y, Humbert-Vidan L, Maule L, Wahid KA, et al. Artificial intelligence and machine learning in cancer pain: a systematic review. J Pain Symptom Manage 2024;68(6):e462-e490 [FREE Full text] [doi: 10.1016/j.jpainsymman.2024.07.025] [Medline: 39097246]

XSL•FO
RenderX

25. Lazris D, Schenker Y, Thomas TH. AI-generated content in cancer symptom management: a comparative analysis between ChatGPT and NCCN. J Pain Symptom Manage 2024;68(4):e303-e311 [FREE Full text] [doi: 10.1016/j.jpainsymman.2024.06.019] [Medline: 38942093]
26. Ott T, Heckel M, Öhl N, Steigleder T, Albrecht NC, Ostgathe C, et al. Palliative care and new technologies. The use of smart sensor technologies and its impact on the total care principle. BMC Palliat Care 2023;22(1):50 [FREE Full text] [doi: 10.1186/s12904-023-01174-9] [Medline: 37101258]
27. Deutsch TM, Pfob A, Brusniak K, Riedel F, Bauer A, Dijkstra T, et al. Machine learning and patient-reported outcomes for longitudinal monitoring of disease progression in metastatic breast cancer: a multicenter, retrospective analysis. Eur J Cancer 2023;188:111-121. [doi: 10.1016/j.ejca.2023.04.019] [Medline: 37229835]
28. Yang TY, Kuo P, Huang Y, Lin H, Malwade S, Lu L, et al. Deep-learning approach to predict survival outcomes using wearable actigraphy device among end-stage cancer patients. Front Public Health 2021;9:730150 [FREE Full text] [doi: 10.3389/fpubh.2021.730150] [Medline: 34957004]
29. Gondode PG, Mahor V, Rani D, Ramkumar R, Yadav P. Debunking palliative care myths: assessing the performance of artificial intelligence chatbots (ChatGPT vs. Google Gemini). IJPC 2024;30:284-287 [FREE Full text] [doi: 10.25259/ijpc_44_2024]
30. Srivastava R, Srivastava S. Can artificial intelligence aid communication? Considering the possibilities of GPT-3 in Palliative care. IJPC 2023;29:418-425 [FREE Full text] [doi: 10.25259/ijpc_155_2023]
31. Reason T, Rawlinson W, Langham J, Gimblett A, Malcolm B, Klijn S. Artificial intelligence to automate health economic modelling: a case study to evaluate the potential application of large language models. Pharmacoecon Open 2024;8(2):191-203 [FREE Full text] [doi: 10.1007/s41669-024-00477-8] [Medline: 38340276]
32. Kamdar M, Lakin J, Zhang H. Artificial intelligence, machine learning, and digital therapeutics in palliative care and hospice: the future of compassionate care or rise of the robots? (TH363). J Pain Symptom Manage 2020;59(2):434-435 [FREE Full text] [doi: 10.1016/j.jpainsymman.2019.12.091]
33. See KC. Using artificial intelligence as an ethics advisor. Ann Acad Med Singap 2024;53(7):454-455 [FREE Full text] [doi: 10.47102/annals-acadmedsg.202471] [Medline: 39132962]
34. Adegbesan A, Akingbola A, Ojo O, Jessica OU, Alao UH, Shagaya U, et al. Ethical challenges in the integration of artificial intelligence in palliative care. J Med, Surg, Public Health 2024;4:100158 [FREE Full text] [doi: 10.1016/j.glmedi.2024.100158]
35. Ranard BL, Park S, Jia Y, Zhang Y, Alwan F, Celi LA, et al. Minimizing bias when using artificial intelligence in critical care medicine. J Crit Care 2024;82:154796 [FREE Full text] [doi: 10.1016/j.jcrc.2024.154796] [Medline: 38552451]
36. De Panfilis L, Peruselli C, Tanzi S, Botrugno C. AI-based clinical decision-making systems in palliative medicine: ethical challenges. BMJ Support Palliat Care 2023;13(2):183-189. [doi: 10.1136/bmjspcare-2021-002948] [Medline: 34257065]
37. Meier LJ, Hein A, Diepold K, Buyx A. Algorithms for ethical decision-making in the clinic: a proof of concept. Am J Bioeth 2022;22(7):4-20. [doi: 10.1080/15265161.2022.2040647] [Medline: 35293841]
38. Macheka S, Ng PY, Ginsburg O, Hope A, Sullivan R, Aggarwal A. Prospective evaluation of artificial intelligence (AI) applications for use in cancer pathways following diagnosis: a systematic review. BMJ Oncol 2024;3(1):e000255. [doi: 10.1136/bmjonc-2023-000255] [Medline: 39886134]
39. Reddy V, Nafees A, Raman S. Recent advances in artificial intelligence applications for supportive and palliative care in cancer patients. Curr Opin Support Palliat Care 2023;17(2):125-134. [doi: 10.1097/SPC.0000000000000645] [Medline: 37039590]
40. Barry C, Paes P, Noble S, Davies A. Challenges to delivering evidence-based palliative medicine. Clin Med (Lond) 2023;23(2):182-184 [FREE Full text] [doi: 10.7861/clinmed.2022-0336] [Medline: 36806204]
41. Chua IS, Gaziel-Yablowitz M, Korach ZT, Kehl KL, Levitan NA, Arriaga YE, et al. Artificial intelligence in oncology: path to implementation. Cancer Med 2021;10(12):4138-4149. [doi: 10.1002/cam4.3935] [Medline: 33960708]
42. Hawker S, Payne S, Kerr C, Hardey M, Powell J. Appraising the evidence: reviewing disparate data systematically. Qual Health Res 2002;12(9):1284-1299. [doi: 10.1177/1049732302238251] [Medline: 12448672]
43. Cavaciuti M, Nwosu AC. 92 ethical challenges of artificial intelligence technology in palliative care. BMJ Supportive Palliative Care 2020;10:A41 [FREE Full text] [doi: 10.1136/spcare-2020-pcc.112]
44. Bak M, Hartman L, Graafland C, Korfage IJ, Buyx A, Schermer M, 4D PICTURE Consortium. Ethical design of data-driven decision support tools for improving cancer care: embedded ethics review of the 4D PICTURE project. JMIR Cancer 2025;11:e65566 [FREE Full text] [doi: 10.2196/65566] [Medline: 40209225]
45. Chow JCL, Li K. Ethical considerations in human-centered AI: advancing oncology chatbots through large language models. JMIR Bioinform Biotechnol 2024;5:e64406 [FREE Full text] [doi: 10.2196/64406] [Medline: 39321336]
46. Johnson JL, Adkins D, Chauvin S. A review of the quality indicators of rigor in qualitative research. Am J Pharm Educ 2020;84(1):7120 [FREE Full text] [doi: 10.5688/ajpe7120] [Medline: 32292186]
47. Karatzanou N. Artificial intelligence (AI) in palliative care: ethical challenges. Bioethica 2025;11(1):51-63 [FREE Full text] [doi: 10.12681/bioeth.40931]
48. Oh O, Demiris G, Ulrich CM. The ethical dimensions of utilizing artificial intelligence in palliative care. Nurs Ethics 2024:09697330241296874 [FREE Full text] [doi: 10.1177/09697330241296874] [Medline: 39551621]

XSL•FO
RenderX

49. Agency for Healthcare Research and Quality. Six domains of healthcare quality [Internet]. URL: https://www.ahrq.gov/talkingquality/measures/six-domains.html [accessed 2025-05-03]
50. Corti C, Cobanaj M, Marian F, Dee EC, Lloyd MR, Marcu S, et al. Artificial intelligence for prediction of treatment outcomes in breast cancer: systematic review of design, reporting standards, and bias. Cancer Treat Rev 2022;108:102410. [doi: 10.1016/j.ctrv.2022.102410] [Medline: 35609495]
51. Zhuang Q, Zhang AY, Cong RSTY, Yang GM, Neo PSH, Tan DS, et al. Towards proactive palliative care in oncology: developing an explainable EHR-based machine learning model for mortality risk prediction. BMC Palliat Care 2024;23(1):124 [FREE Full text] [doi: 10.1186/s12904-024-01457-9] [Medline: 38769564]
52. Bowers A, Drake C, Makarkin AE, Monzyk R, Maity B, Telle A. Predicting patient mortality for earlier palliative care identification in medicare advantage plans: features of a machine learning model. JMIR AI 2023;2:e42253 [FREE Full text] [doi: 10.2196/42253] [Medline: 38875557]
53. Ortiz BL, Gupta V, Kumar R, Jalin A, Cao X, Ziegenbein C, et al. Data preprocessing techniques for AI and machine learning readiness: scoping review of wearable sensor data in cancer care. JMIR mHealth uHealth 2024;12:e59587 [FREE Full text] [doi: 10.2196/59587] [Medline: 38626290]
54. Kalla M, O'Brien T, Metcalf O, Hoda R, Chen X, Li A, et al. Understanding experiences of telehealth in palliative care: photo interview study. JMIR Hum Factors 2025;12:e53913 [FREE Full text] [doi: 10.2196/53913] [Medline: 39935022]
55. Bertsimas D, Dunn J, Pawlowski C, Silberholz J, Weinstein A, Zhuo YD, et al. Applied informatics decision support tool for mortality predictions in patients with cancer. JCO Clin Cancer Inform 2018;2:1-11 [FREE Full text] [doi: 10.1200/CCI.18.00003] [Medline: 30652575]
56. Roldan-Vasquez E, Mitri S, Bhasin S, Bharani T, Capasso K, Haslinger M, et al. Reliability of artificial intelligence chatbot responses to frequently asked questions in breast surgical oncology. J Surg Oncol 2024;130(2):188-203. [doi: 10.1002/jso.27715] [Medline: 38837375]

Abbreviations

AI: artificial intelligence
CASP: Critical Appraisal Skills Programme
IOM: Institute of Medicine
ML: machine learning
OECD: Organisation for Economic Co-Operation and Development
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SDOH: social determinants of health

Edited by T de Azevedo Cardoso; submitted 06.03.25; peer-reviewed by E Ray Chaudhuri, MK Milic; comments to author 14.04.25; revised version received 22.04.25; accepted 29.04.25; published 14.05.25.

Please cite as:
García Abejas A, Geraldes Santos D, Leite Costa F, Cordero Botejara A, Mota-Filipe H, Salvador Vergés À
Ethical Challenges and Opportunities of AI in End-of-Life Palliative Care: Integrative Review
Interact J Med Res 2025;14:e73517
URL: https://www.i-jmr.org/2025/1/e73517
doi:10.2196/73517
PMID:40302210

©Abel García Abejas, David Geraldes Santos, Fabio Leite Costa, Aida Cordero Botejara, Helder Mota-Filipe, Àngels Salvador Vergés. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.05.2025. 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

Guideline-Based Digital Exercise Interventions for Reducing Body Weight and Fat and Promoting Physical Activity in Adults With Overweight and Obesity: Systematic Review and Meta-Analysis

Mohamad Motevalli[1,2], MSc; Clemens Drenowatz[3], Prof Dr; Derrick Tanous[1,2], MSc; Gerhard Ruedl[1], Prof Dr; Werner Kirschner[1], Prof Dr; Markus Schauer[1], MSc; Thomas Rosemann[4], Prof Dr; Katharina Wirnitzer[1,2,5,6], Prof Dr

[1]Department of Sport Science, University of Innsbruck, Innsbruck, Austria

[2]Department of Secondary Education, University College of Teacher Education Tyrol, Innsbruck, Austria

[3]Division of Sport, Physical Activity and Health, University of Teacher Education Upper Austria, Linz, Austria

[4]Institute of Primary Care, University of Zurich, Zurich, Switzerland

[5]Department of Pediatric Oncology and Hematology, Charité – University of Medicine Berlin, Berlin, Germany

[6]Charité Competence Center for Traditional and Integrative Medicine, Charité – University of Medicine Berlin, Berlin, Germany

Corresponding Author:
Mohamad Motevalli, MSc
Department of Sport Science
University of Innsbruck
Fürstenweg 185
Innsbruck, 6020
Austria
Phone: 43 67763778121
Email: mohamad.motevalli@uibk.ac.at

Abstract

Background: Digitally delivered physical exercise interventions are becoming increasingly popular in addressing the obesity epidemic. However, there remains uncertainty on their efficacy regarding the reduction of body weight (BW) and body fat, which may, at least partly, be due to variations in study designs and inconsistent adherence to international physical activity (PA) guidelines.

Objective: This study aimed to evaluate the effectiveness of digital exercise interventions based on PA guidelines in reducing BW and fat in adults with overweight or obesity, as well as their impact on PA-related factors.

Methods: This review was conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Comprehensive searches were performed in October 2024 across PubMed, Cochrane Library, Web of Science, and Ovid MEDLINE databases. Eligible studies included adults (aged ≥18 years) with objectively confirmed overweight or obesity who used digital interventions aligned with international PA guidelines. Risk of bias was evaluated using the Cochrane Risk of Bias (version 2) tool for randomized controlled trials and the Risk of Bias in Nonrandomized Studies of Interventions tool for nonrandomized studies. A random-effects meta-analysis with Hartung-Knapp adjustment was performed using R software.

Results: Out of 4948 studies identified, 188 (3.8%) were screened in full and 30 (0.6%) met the eligibility criteria. Intervention durations ranged from 8 weeks to 24 months (average 6.4, SD 5.5 months). Meta-analysis showed that guideline-based digital exercise interventions significantly reduced BW compared to controls (mean difference [MD]=−1.17 kg; P=.003; I^2=0.0%), with subgroup analysis revealing greater effects in active (nondigital) controls (MD=−1.23 kg; I^2=7.5%) compared to passive (waitlist) controls (MD=−0.52 kg; I^2=0.0%). A significant reduction in BMI was observed (MD=−0.50 kg/m^2; P=.003), although with substantial heterogeneity (I^2=70.0%), and subgroup analysis showed greater effects compared to passive controls (MD=−0.70 kg/m^2; I^2=43.1%) rather than to active controls (MD=−0.45 kg/m^2; I^2=74.5%). No significant effect was observed for body fat percentage overall (MD=−0.08%; P=.84; I^2=7.4%). Qualitative analysis (including findings from noncomparative studies) showed that guideline-based digital exercise interventions led to significant reductions in BW (22/25, 88% studies; range −1.3 to −8.4 kg); BMI (19/23, 83% of studies; range −0.4 to −3.4 kg/m^2); waist circumference (15/16, 94% of studies; range −2.1 to −9.2 cm),

XSL•FO
RenderX

body fat percentage (9/9, 100% of studies; range −0.3% to −4.1%); and fat mass (7/7, 100% of studies; range −0.4 to −6.5 kg), while findings for waist-to-hip ratio and PA outcomes were inconsistent.

Conclusions: Guideline-based digital PA and exercise interventions show potential in reducing excess BW in adults with overweight or obesity, with stronger effects when compared to nondigital interventions. However, their superiority over traditional methods is uncertain for BMI and body composition. Substantial variations in study designs present challenges in drawing definitive conclusions on specific characteristics of effective digital exercise tools.

Trial Registration: PROSPERO CRD42024620020; https://www.crd.york.ac.uk/PROSPERO/view/CRD42024620020

(Interact J Med Res 2025;14:e73656) doi:10.2196/73656

KEYWORDS

technology; mobile health; mHealth; eHealth; ubiquitous Health; uHealth; smartphone; weight loss; body composition; adulthood; training; fitness

Introduction

Background

Overweight and obesity are among the most pressing public health challenges of the 21st century, posing significant health risks and economic burdens worldwide [1]. According to the World Health Organization (WHO), the prevalence of adult obesity has more than doubled since 1990 [2]. By 2022, 43% of adults aged ≥18 years were classified as overweight, and 16% were living with obesity, making it a major risk factor for several chronic diseases, including cardiovascular disease, type 2 diabetes, musculoskeletal disorders, and cancer [2]. Data shows that being overweight or obese is associated with approximately 2.8 million deaths each year [3]. This highlights the urgent need for effective interventions to manage body weight (BW) and improve body composition in adults with excess body fat.

Physical activity (PA) is recognized as a critical component of obesity prevention and management strategies [4]. The WHO guidelines recommend that adults engage in at least 150 to 300 minutes of moderate-intensity aerobic activity or 75 to 150 minutes of vigorous-intensity activity per week [5]. It has been well documented that adherence to these recommendations is associated with significant improvements in body composition [6], emphasizing the critical role of PA in mitigating obesity-related health issues. Despite its proven benefits, adults often struggle to maintain sufficient levels of PA due to barriers such as time constraints, lack of access to fitness facilities, and motivation, highlighting the need for innovative solutions that make PA and exercise more accessible and sustainable [7].

Today's digital world represents a transformative change in how we live, work, and interact [8], characterized by the integration of digital technologies into almost every aspect of human life. One of the most significant outcomes of this technological revolution is the expansion of mobile apps, which have become integral tools in daily life. An app is defined as software designed for specific tasks, which can be downloaded and installed on mobile phones or other digital devices [9]. In recent years, the advancement of digitalization has led to the emergence of digital health interventions as promising tools for supporting BW management and promoting healthy lifestyles [10]. Among these, PA and fitness apps have gained considerable attention by leveraging technological advancements to offer and deliver exercise plans, while also monitoring PA and fitness progress [11].

Over the past 2 decades, several studies have examined the effectiveness of health and fitness apps in BW management and other health outcomes [12-22]; however, the results have been inconsistent, with some studies reporting significant benefits while others found minimal or no effect on target outcomes. This variability in results may be attributed to differences in study design, population characteristics, app features, and intervention characteristics. More importantly, existing studies on digital PA and exercise interventions often lack a standardized implementation of key exercise characteristics (such as type, frequency, duration, and intensity) based on international PA guidelines. This inconsistency makes it difficult to compare interventions and assess their adequacy, thereby potentially influencing the outcomes. Therefore, a comprehensive review of digital PA interventions that are based on PA guidelines, such as those from the WHO or an equivalent framework, is needed to ensure more consistent and reliable outcomes.

Objectives

The primary objective of this systematic review was to evaluate the effectiveness of digital exercise interventions that were based on international PA recommendations on body composition, as well as PA and fitness outcomes, in adults with overweight and obesity. In addition, while this review aimed to identify gaps in the literature, its overarching goals were to provide insights that inform clinical practice and public health initiatives, and to guide the design of future digital interventions, thereby contributing to the effective management of overweight and obesity on a broader scale.

Methods

Study Protocol

This systematic review and meta-analysis was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [23]. The PRISMA checklist was used to guide the reporting across all stages of the review, encompassing the formulation of the research question, establishment of eligibility criteria, development of the search strategy, study selection, data extraction, and synthesis of results. The review protocol was

XSL•FO
RenderX

registered prospectively in the PROSPERO under the registration number CRD42024620020. The protocol's title was revised during the review process; however, no other deviations from the registered protocol occurred.

Eligibility Criteria

The eligibility criteria encompassed multiple dimensions guided by the population, intervention, comparison, outcome, and study design framework [24], along with additional considerations. The population of interest comprised adults aged ≥18 years with overweight or obesity. Overweight and obesity were defined using established thresholds: a BMI ≥25.0 kg/m^2 [25] or ≥23.0 kg/m^2 for Asia-Pacific adult populations [26], body fat percentage (BF%) ≥25% for male individuals and ≥36% for female individuals [27], or a waist-to-hip ratio (WHR) ≥0.90 for male individuals and ≥0.80 for female individuals [28]. Baseline measures of BW status had to be objectively assessed; studies relying on self-reported anthropometric data were excluded. To ensure the feasibility of participation in PA and exercise interventions, studies were required to include participants without severe health conditions or physical limitations that could hinder PA and exercise engagement. Studies involving trained participants who had been regularly engaged in PA, exercise, or sports within the 3 months before the intervention were also excluded. Furthermore, studies were excluded if they included participants who were pregnant or within 1 year postpartum, as these conditions can significantly influence body composition.

Regarding the intervention criteria, the review targeted interventional studies assessing the effectiveness of digital PA and exercise programs designed in alignment with WHO guidelines on PA, that is, a minimum of 150 minutes of moderate-intensity PA or 75 minutes of vigorous-intensity PA per week [5], or a comparable framework. "PA" refers to any bodily movement produced by skeletal muscles that requires energy expenditure, while "exercise" or "physical exercise" is a subset of PA that is planned, structured, and repetitive, with the specific objective of achieving or maintaining health and fitness [29,30]. Eligible studies implemented PA or exercise interventions with a minimum duration of 8 weeks. In controlled trials, the control group (whether consisting of participants not engaging in physical exercise or those who did not engage in any digital interventions) was considered the comparator. For single-arm interventional studies, comparisons were made against baseline values of the same participants assessed before the intervention.

The primary outcomes of interest included anthropometric and body composition measures: BW, BMI, BF%, fat mass (FM), waist circumference (WC), and WHR. All outcomes had to be objectively measured; studies relying on self-reported data were excluded to minimize bias and ensure reliability. Any study that reported at least one of these outcomes was eligible for inclusion. The secondary outcomes included PA-related or physical fitness variables, regardless of the assessment method. Finally, eligible studies were restricted to publications in English or German, and no limitations were applied regarding the publication period, context, setting, or geographic location.

Information Sources and Search Strategy

A comprehensive search strategy was used to identify relevant studies for inclusion in this systematic review, conducted from October 1 to 10, 2024. The databases searched included PubMed, Cochrane Library, Web of Science, and Ovid MEDLINE. The search strategy used a combination of keywords and Medical Subject Headings (MeSH) terms with the Boolean operators "AND" and "OR," as detailed in Multimedia Appendix 1. In addition to the database searches, the reference lists of the included studies and relevant review articles were manually reviewed to identify any additional publications that may not have been captured in the database search.

Study Selection and Data Extraction

The study selection process was conducted using Covidence, a specialized tool for systematic screening and data extraction. Two independent reviewers (MM and DT) screened the titles and abstracts of all retrieved articles based on the predefined eligibility criteria. Any discrepancies between the 2 reviewers were resolved through discussion. Articles deemed potentially relevant during the abstract screening phase underwent full-text review, and the same reviewers thoroughly assessed the full text of each article, while reasons for exclusion were documented. A third reviewer (CD) was consulted for the final decision if consensus could not be reached. Studies that met all the eligibility criteria during the full-text screening process advanced to the data extraction phase, where the same reviewers extracted the data. The extracted data encompassed several domains, including study characteristics, study design, intervention details, sample size, population, app or digital tool specifics, PA or exercise protocols, comparators, anthropometric and body composition outcomes, and PA and fitness outcomes. When the data were unclear, incomplete, or missing, the reviewers contacted the authors of the studies to request clarifications or additional information. When multiple articles reported the same study with identical population, outcomes, and data, the earliest publication was included to avoid data duplication. Any exclusions were carefully considered and documented to ensure transparency in the selection process.

Risk of Bias Assessment

The risk of bias in the included studies was assessed using standardized tools to ensure a thorough evaluation of methodological quality. Two independent reviewers (MM and DT) assessed the risk of bias, and any discrepancies were resolved through discussion. For randomized controlled trials (RCTs), the Cochrane Risk of Bias (version 2) tool was used [31]. This tool evaluates bias across 5 key domains: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. Each domain is assessed for potential sources of bias that could impact the validity of the trial's findings. The Risk of Bias in Nonrandomized Studies of Interventions tool was used for nonrandomized and single-arm studies [32]. This tool evaluates 7 domains: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, outcome measurement, and selection of the reported result. Using both tools enabled a comprehensive assessment of the methodological

XSL•FO
RenderX

rigor and potential biases in each study, ensuring a reliable interpretation of the evidence.

Data Synthesis

A qualitative synthesis of the data was conducted, focusing on the various outcomes reported across studies. The extracted data were systematically grouped and analyzed by outcome categories and variables (study design, population, study group, app or digital tool, intervention, primary outcomes, and PA outcomes), enabling a detailed exploration of trends and patterns in the evidence. The synthesis aimed to identify consistent trends across studies while also highlighting areas of agreement, inconsistencies, or discrepancies in the findings. This approach provided a deeper understanding of the broader implications of the data and helped identify areas where further research is needed to resolve conflicting results or clarify uncertainties.

Meta-analysis was performed using R (version 4.5.0; R Foundation for Statistical Computing) to evaluate the effects of digital exercise interventions on primary outcome variables. A random-effects model with Hartung-Knapp adjustment was applied to account for potential between-study heterogeneity and to provide robust CIs. The primary effect size metric was the mean difference (MD) between intervention and control groups, expressed in the original units for each outcome. Subgroup analyses were conducted to examine potential differences in intervention effects by control type. Studies were stratified based on the type of control condition: (1) active controls (ie, nondigital interventions or usual care) and (2) passive controls (ie, waitlist—participants assigned to receive the intervention after the study period—or no intervention). When studies included multiple eligible intervention arms, each arm was treated as a separate unit of analysis to avoid statistical dependency, with the intervention arm serving as the unit of analysis. Heterogeneity was assessed using the I^2 statistic, τ^2, and Cochran Q test. All procedures were conducted following the PRISMA guidelines, and results are reported with 95% CIs and corresponding *P* values.

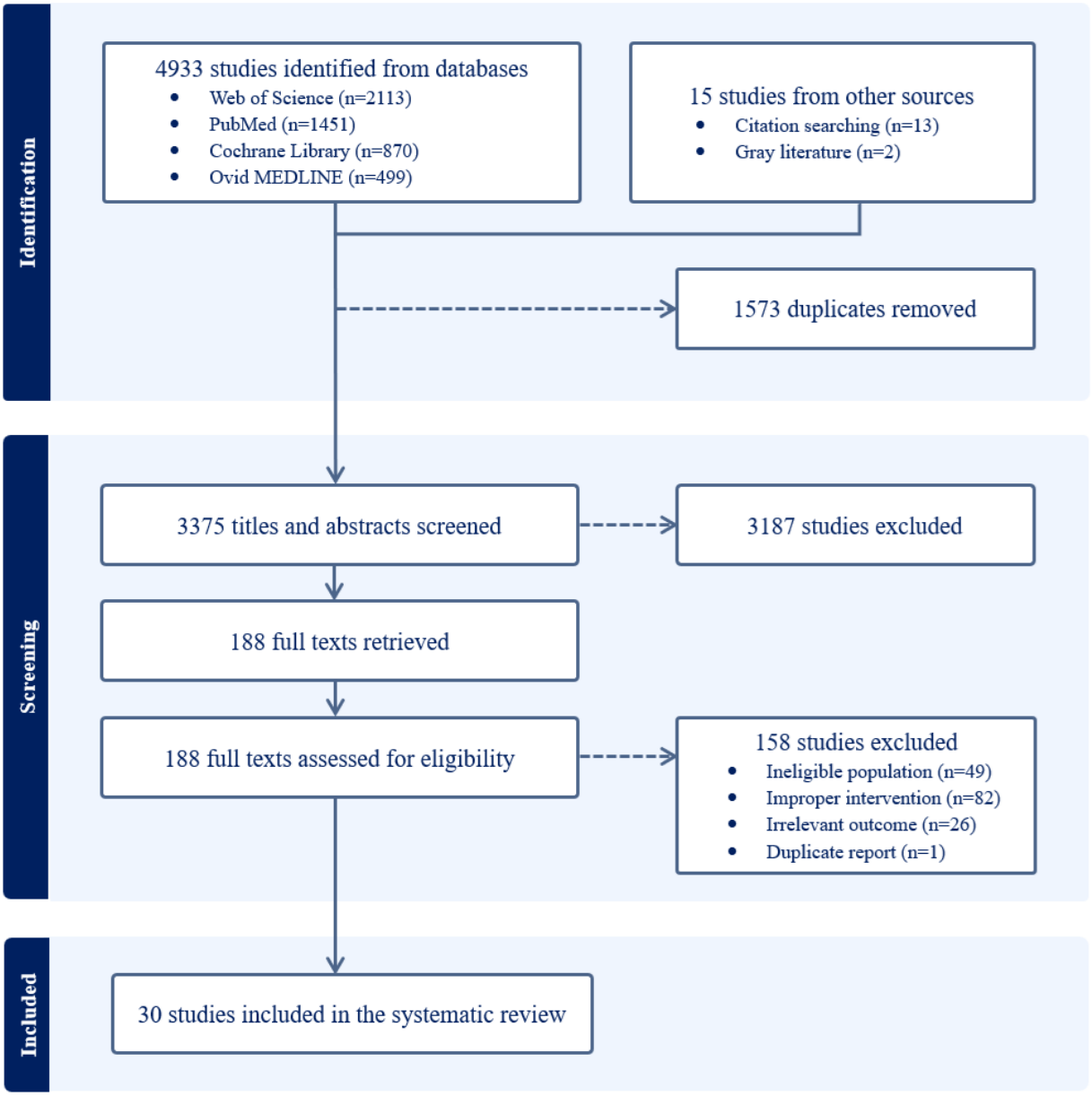
Results

Selection Process

A total of 4948 studies were initially identified through the databases and reference sources. After removing duplicates, 68.21% (3375/4948) of the studies were screened by title and abstract, of which 5.57% (188/3375) of the studies were retrieved for full-text review. Ultimately, 30 studies met the predefined eligibility criteria and were included in the systematic review. Figure 1 presents the PRISMA flowchart outlining the study selection process.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart showing the study identification, screening, and inclusion process.



Included Studies

The final sample of included studies consisted of 26 RCTs and 4 nonrandomized studies. Out of the 30 included studies, 11 (37%) were conducted in the United States, 3 (10%) in Australia, 3 (10%) in Korea, 2 (7%) in the United Kingdom, 2 (7%) in Germany, 2 (7%) in Spain, and 1 (3%) each in Sweden, France, Canada, Belgium, Singapore, China, and Hong Kong. The sample sizes varied from 16 to 750 participants, with an average of 132 participants. Of the 30 included studies, 6 (20%) exclusively included female participants, while none of the included studies focused solely on male participants. Table 1 provides a summary of the characteristics of the studies included in the review.

XSL•FO
RenderX

Table 1. Summary of the characteristics and outcomes of the studies included in the systematic review. The studies are ordered by intervention duration.

Study, year, and country	Design and population	Study groups	Intervention	Primary outcomes and findings	PA[a] outcomes and findings
Intervention duration: 24 months					
Jakicic et al [33], 2016, United States	Two-arm RCT[b]; sample size: 470; median age 30.9 years; 71.1% female participants; median BMI 31.2 kg/m^2	• Enhanced intervention (n=237) • Standard intervention (n=233)	Digital tool: FIT Core; BodyMedia; a 24-month intervention where all participants initially increased PA, followed a low-calorie diet, and attended group counseling. At 6 months, telephone counseling, text message prompts, and website access were introduced. The standard group used a website for self-monitoring, while the enhanced group used a wearable device with a web interface. Moderate to vigorous PA progressed from 100 to 300 min per week over 4-week intervals.	Significant decreases in BW[c] BMI, BF%[d], and FM[e] in both groups; no significant between-group differences for BMI, BF%, and FM, while postintervention BW was significantly lower in the standard group.	Significant improvements in LPA[f] and MVPA[g] in both groups; no significant between-group differences
Svetkey et al [34], 2015, United States	Three-arm RCT; sample size: 365; mean age 29.4 years; 69.6% female participants; mean BMI 35.2 kg/m^2	• mHealthh (n=123) • Personal coaching (n=120) • Control (n=122)	Digital tool: Investigator-designed app; a 24-month intervention delivered via a smartphone app based on social cognitive theory and transtheoretical model, incorporating goal setting, behavioral self-management, and motivational enhancement, with a PA target of achieving ≥180 min per week of moderate PA.	Significant differences in BW between the study groups, with greater reductions in the personal coaching group at 6 and 12 months, but not at 24 months.	—[i]
Intervention duration: 12 months					
Héroux et al [35], 2017, Canada	Single-arm intervention trial; sample size: 28; mean age 49.6 years; 100% female participants; mean BMI 32.2 kg/m^2	• Intervention (n=28)	Digital tool: The online Precision Nutrition Coaching Program; a 12-month online program focused on daily physical exercise, dietary habits, and health behavior lessons, with coaching delivered via computer, tablet, and mobile device. The exercise program included detailed daily workouts (aerobic interval training, weight training, or walking) with progressive intensity over time.	Significant decreases in BW, WC[j], and FM (but not in BMI).	—
Lynch et al [36], 2017, United States	Single-arm intervention trial; sample size: 46; mean age 39 years; 65.2% female participants; mean BMI 31.8 kg/m^2	• Intervention (n=46)	Digital tool: Lean Eating by Precision Nutrition Coaching; a 12-month weight loss intervention with daily components of physical exercise, nutrition, behavioral modification strategies, and health lessons via an online platform, along with online coaching. The PA component included at least 150 min of physical exercise per week, with a minimum of 2 strength training sessions per week.	Significant decreases in BW, BMI, and BF% (but not in WHR[k]).	—
Watson et al [37], 2015, United Kingdom	Two-arm RCT; sample size: 65; mean age 52.2 years; 55.4% female participants; mean BMI 32.6 kg/m^2	• Web-based (n=32) • Usual self-care (n=33)	Digital tool: Imperative Health; a 12-month web-based weight loss intervention incorporating PA and dietary components, along with goal setting and tailored feedback and support provided by physiologists via telephone and email. The PA program included a weekly schedule for planning activities at light to vigorous intensity levels.	Significant decreases in BW in the intervention group, with significant between-group differences at 3 and 6 months, but not at 12 months. Significant between-group differences in BMI and WC at 3 and 6 (but not at 12) months.	Significant between-group differences in daily PA at 3 months, but not at 6 and 12 months.
Intervention duration: 6 months					

XSL•FO
RenderX

Study, year, and country	Design and population	Study groups	Intervention	Primary outcomes and findings	PA[a] outcomes and findings
Alcántara-Aragón et al [38], 2018, Spain	Two-arm RCT; sample size: 183; mean age 44.5 years; 83.6% female participants; mean BMI 34.75 kg/m^2	• Telematic intervention (n=91) • Nontelematic intervention (n=92)	Digital tool: PREDIRCAM2 web platform; a 6-month tailored physical exercise and dietary plans based on international PA guidelines and the Mediterranean diet. The telematic group received online support and virtual contact, while the nontelematic group followed traditional methods.	Significant decreases in BW, WC, and WHR in the telematic intervention group; no significant between-group differences.	—
Allen et al [39], 2013, United States	Four-arm RCT; sample size: 68; mean age 44.9 years; 77.9% female participants; mean BMI 34.3 kg/m^2	• Intensive counseling (n=18) • Smartphone only (n=17) • Intensive counseling+smartphone (n=16) • Less intensive counseling+smartphone (n=17)	Digital tool: Lose It!; a 6-month intervention (PA and diet) using an eclectic approach combining social cognitive theory, behavioral self-management, and motivational interviewing techniques to achieve 5% BW loss and a minimum of 150 minutes of moderate-intensity PA per week. Intensive groups attended frequent in-person sessions, while less intensive and smartphone-based groups had fewer sessions.	Significant decreases in BW, BMI, and WC in all groups; no significant between-group differences.	No significant between-group difference in MVPA.
Apiñaniz et al [40], 2019, Spain	Two-arm RCT; sample size: 110; mean age 38.5 years; 71.8% female participants; mean BMI 32.7 kg/m^2	• Health advice+app (n=54) • Health advice (n=56)	Digital tool: AKTIDIET; a 6-month intervention consisting of health advice (PA and dietary) based on international guidelines for both groups. In the intervention group, this advice was reinforced through an app consisting of aerobic and muscle training programs and food intake tracking.	Significant decreases in BW in both groups; no significant between-group differences.	—
Batsis et al [41], 2021, United States	Single-arm intervention trial; sample size: 53; mean age 72.9 years; 69.8% female participants; mean BMI 36.5 kg/m^2	• Intervention (n=53)	Digital tool: Video conferencing; a 6-month weight management program combining synchronous videoconference-based physical exercise and nutrition sessions, remote Fitbit monitoring, and periodic face-to-face interactions. Participants attended 75-minute, twice-weekly group exercise sessions led by a trained physical therapist.	Significant decreases in BW, BMI, and WC; no change in WHR.	Significant improvements in 30-second sit-to-stand and 6-minute walk test; no changes in gait speed or grip strength.
Block et al [42], 2015, United States	Two-arm RCT; sample size: 339; mean age 55.0 years; 31.3% female participants; mean BMI 31.2 kg/m^2	• Alive-PD intervention (n=163) • Usual care control (n=176)	Digital tool: Alive-PD; a 6-month program with weekly goal setting (PA and dietary) delivered via web, email, interactive voice response calls, and a supportive mobile app. The PA target included 150-300 minutes of aerobic activity and resistance training weekly, based on baseline levels and progress.	Significant decreases in BW, BMI, and WC in both groups, with significantly greater reductions in the intervention group.	—
Duncan et al [18], 2020, Australia	Three-arm RCT; sample size: 116; mean age 44.5 years; 70.7% female participants; mean BMI 31.7 kg/m^2	• Enhanced (n=39) • Traditional (n=41) • Waitlist control (n=36)	Digital tool: Balanced app; a 6-month multicomponent (PA and dietary) mHealth intervention delivered via a smartphone app, providing educational materials, goal setting, self-monitoring, and feedback; the enhanced group received additional sleep-related support. PA action planning focused on MVPA, resistance training, and step count increases.	No significant between-group difference in BW, while postintervention WC was significantly lower in the traditional group.	No significant between-group differences in MVPA and LPA.
Hartman et al [43], 2016, United States	Two-arm RCT; sample size: 55; mean age 59.5 years; 100% female participants; mean BMI 31.9 kg/m^2	• Technology-based (n=36) • Usual care (n=18)	Digital tool: Fitbit and a website; a 6-month weight loss intervention focused on developing and practicing self-monitoring and self-regulatory skills, with a goal of losing 10% of BW by engaging in at least 150 minutes per week of MVPA and restricting calorie intake.	Significant decreases in BW in the intervention group, with a significant between-group difference (no BW change in the usual care group).	Significant improvement in daily MVPA in the intervention group, with no significant between-group differences.

XSL•FO
RenderX

Study, year, and country	Design and population	Study groups	Intervention	Primary outcomes and findings	PA[a] outcomes and findings
Haufe et al [44], 2021, Germany	Two-arm RCT; sample size: 314; mean age 48.1 years; 14% female participants; mean BMI 33.3 kg/m^2	• Telemonitoring physical exercise (n=160) • Waitlist control (n=154)	Digital tool: Custom-designed smartphone app; a 6-month telemonitoring-supported lifestyle intervention focused on regular physical exercise, including individual exercise recommendations provided during face-to-face meetings and via a smartphone app, aiming for at least 150 minutes of PA per week.	Significant decrease in WC in the intervention group, with a significant between-group difference (less but significant decrease in WC in the control group).	Significant increase in exercise capacity in both groups, with a significantly greater increase in the intervention group.
Hutchesson et al [20], 2018, Australia	Two-arm RCT; sample size 57; mean age 27.1 years; 100% female participants; mean BMI 29.4 kg/m^2	• Intervention (n=29) • Control (n=28)	Digital tool: Be Positive Be Health e; a 6-month weight loss program delivered via eHealth technologies (website, app, email, text messages, and social media), incorporating goal setting and grounded in social cognitive theory and control theory, aligned with international PA and dietary guidelines.	Significant decreases in BW, BMI, FM, and BF% in the intervention group, with no significant between-group differences except for FM. Significant decreases in WC in both groups.	No significant changes in MVPA, MPA[l], and VPA[m] in any study group.
Lim et al [45], 2022, Singapore	Two-arm RCT; sample size: 148; mean age 53.1 years; 39.9% female participants; mean BMI 29.8 kg/m^2	• Intervention (n=72) • Control (n=76)	Digital tool: nBuddy Diabetes app; a 6-month in-app dietitian coaching incorporating behavioral strategies (goal setting, stimulus control, problem-solving, self-monitoring, cognitive restructuring, and motivational interviewing), with a PA goal of 150 minutes per week of moderate-intensity physical exercise.	Significant decreases in BW and BMI in both groups, with significantly greater reductions in the intervention group.	Significant improvement in general PA in the intervention group; no significant between-group difference.
Rogers et al [46], 2016, United States	Three-arm RCT; sample size: 39; mean age 39.9 years; 79.5% female participants; mean BMI 39.5 kg/m^2	• Standard (n=14) • Technology (n=12) • Enhanced technology (n=13)	Digital tool: BodyMedia Fit and LINK; a 6-month mHealth intervention program incorporating PA and dietary components, alongside self-monitoring and feedback. The PA component involved unsupervised, home-based physical exercise at moderate intensity, with a gradual increase in session duration.	Significant decreases in BW, BMI, WC, BF%, and FM in all groups; no significant between-group differences.	—
Intervention duration: 12-16 weeks					

XSL•FO
RenderX

Study, year, and country	Design and population	Study groups	Intervention	Primary outcomes and findings	PA[a] outcomes and findings
Hurst et al [47], 2021, United States	Single-arm intervention trial; sample size: 30; mean age 38.1 years; 43% female participants; median BMI 32.7 kg/m^2	• Intervention (n=30)	Digital tool: Telehealth coaching; 16-week telehealth intervention delivered in primary care clinics, using wearable devices, automated text messaging, and trained health coaching (including real-time feedback) to support physical exercise and nutrition goals, including achieving at least 150 minutes of PA per week.	Significant decreases in BW and BMI.	Significant improvement in MVPA.
Stephens et al [48], 2017, United States	Two-arm RCT; sample size: 62; median age 20.0 years; 71% female participants; median BMI 28.5 kg/m^2	• Smartphone + health coach (n=31) • Control (n=31)	Digital tool: Lose it!; a 3-month intervention using a behavior-based smartphone app (primarily grounded in self-efficacy theory) for weight loss, combined with counseling sessions and text messages from a health coach, with goals of losing 1-2 pounds per week and engaging in at least 150 minutes of moderate-intensity PA weekly.	Significant decreases in BW, BMI, and WC in the intervention group (but not in the control group), with significant between-group differences.	Significant improvement in PA score in the intervention group, with no significant between-group difference.
Wong et al [49], 2021, Hong Kong	Two-arm RCT; sample size: 77; mean age 58.9 years; 55.8% female participants; mean BMI 27.02 kg/m^2	• App (n=38) • Booklet (n=39)	Digital tool: MetS app; a 3-month lifestyle intervention delivered via a mobile app to support participants' individual physical exercise routines through goal setting, logging exercise type and duration, and self-monitoring. The PA goal was at least 30 minutes per day, 5 days per week.	Significant decreases in BW and BMI (but not in WC or WHR) in the app group, with significant between-group differences for BW and BMI (no change in the booklet group).	Significant improvements in exercise score and exercise self-efficacy in the app group, with significant between-group differences.
Zhang et al [50], 2023, China	Three-arm RCT; sample size: 750; mean age 70.1 years; 53.9% female participants; mean BMI 27.7 kg/m^2	• Remote PA and diet (n=250) • Remote PA (n=250) • Control (n=250)	Digital tool: mHealth app; a 3-month remote weight management intervention focused on PA (with or without dietary components) including health assessments and guidance from exercise instructors and nutritional professionals. The PA program included 20 minutes of resistance or aerobic exercise or walking 6000 steps daily.	Significant decreases in BW, BMI, WC, and WHR in the remote PA and diet group, with significant between-group differences at day 90, but not at day 45.	No significant changes in total PA and PA levels in any study group.
Anderson et al [51], 2018, United Kingdom	Two-arm RCT; sample size: 78; mean age 47.1 years; 88% female participants; mean BMI 32.7 kg/m^2	• Intervention (n=39) • Control: lifestyle booklet (n=39)	Digital tool: LivingWELL; a 12-week intervention consisting of 1 face-to-face session, 4 phone consultations, and web-based support to achieve 5% BW reduction. It included a personalized diet and physical exercise plan with behavioral techniques such as motivational interviewing and action planning.	Decreases in BW, BMI, and WC in the intervention group, but not in the control group.	Increases in daily MPA and daily steps (but not daily VPA) in the intervention group.
Bughin et al [52], 2021, France	Two-arm RCT; sample size: 50; mean age 52.2 years; 54% female participants; mean BMI 36.5 kg/m^2	• Telerehabilitation program (n=25) • Usual care (n=25)	Digital tool: Telemouv app; a 12-week multicomponent intervention (PA, diet, and education) available on smartphones and a website. The PA program included endurance exercises (targeting 150 minutes of PA per week), muscle strengthening (with gradual increases in volume), and balance exercises.	Significant decreases in BF% and FM in both groups, with no significant changes in BW, BMI, or WHR. No significant between-group differences for any outcomes.	No significant within- or between-group change or difference in the 6-minute walk test.
Hebden et al [53], 2014, Australia	Two-arm RCT; sample size: 51; mean age 22.8 years; 81% female participants; mean BMI 27.3 kg/m^2	• mHealth intervention (n=26) • Control (n=25)	Digital tool: ePASS; a 12-week mHealth program targeting key lifestyle behaviors associated with weight gain, delivered via smartphone apps, internet forums, text messages, and emails. The PA program included 30 minutes of moderate-intensity PA per day for general health and 60 minutes per day for weight management.	Significant decreases in BW and BMI in both groups; no significant between-group differences.	Significant increase in daily LPA (but not MVPA) in the intervention group.

XSL•FO
RenderX

Study, year, and country	Design and population	Study groups	Intervention	Primary outcomes and findings	PA[a] outcomes and findings
Hong et al [54], 2022, Korea	Two-arm RCT; sample size: 31; mean age 79.9 years; 100% female participants; mean BF% 41.6	• Intervention (n=12) • Control (n=17)	Digital tool: Smartphone telepresence platform; a 12-week smartphone mirroring-based telepresence physical exercise program conducted at home with 3 sessions per week, primarily focused on resistance training (with gradual intensity increases), all based on international guidelines; while the control group performed the same program in person.	Significant decreases in BF% in both groups and in BW in the control group; no significant between-group differences.	Significant left-hand grip strength improvement in the intervention group and right-hand in the controls; no significant between-group differences.
Hurkmans et al [16], 2018, Belgium	Four-arm RCT; sample size: 81; mean age 45.0 years; 71.6% female participants; mean BMI 32.0 kg/m^2	• Face-to-face (n=28) • App (n=30) • Combined (n=22) • Control (n=22)	Digital tool: Mobile Weight Loss App; 12-week weight loss programs using face-to-face, mobile, and combined approaches, offering physical exercise and dietary guidance aligned with standard guidelines, along with personalized exercise plans provided by a coach.	Significant decreases in BMI in the face-to-face, app, and combined groups; no significant difference between the 3 interventions.	Significant improvements in MVPA; no significant between-group differences.
Johnson et al [55], 2019, United States	Three-arm RCT; sample size: 30; mean age 43.2 years; mean BMI 36.1 kg/m^2	• Videoconferencing (n=10) • In-person (n=10) • Control (n=10)	Digital tool: Withings app and Healow app; 12-week telemedicine-based health coaching, including physical exercise routines, goal setting, and PA progression, with feedback provided by a registered dietitian and exercise physiologist. PA program followed international guidelines recommending at least 150 minutes of PA per week (30 minutes of MVPA, 5 days per week).	Significant differences in BW reduction between the study groups, with greater reduction in the videoconferencing group; no significant between-group differences in BMI. Decrease in BW and BMI in all groups.	Improvement in daily steps in all groups, with a significantly greater improvement in the videoconferencing group.
Pressler et al [56], 2010, Germany	Two-arm RCT; sample size: 105; median age 48 years; 11.8% female participants; mean BMI 29.0 kg/m^2	• Structured web-based exercise (n=66) • Nonstructured web-based exercise (n=39)	Digital tool: Interactive internet platform; a 12-week intervention with structured or nonstructured internet-delivered physical exercise programs performed individually, alongside education sessions. The exercise program included 3 moderate endurance sessions and 1 strength training session per week.	Significant decreases in WC in both groups and in BMI and BF% in the control group; no significant between-group differences in any of these outcomes.	No significant changes in daily step counts in any study group.
Intervention duration: 8-10 weeks					

XSL•FO
RenderX

Study, year, and country	Design and population	Study groups	Intervention	Primary outcomes and findings	PA[a] outcomes and findings
Ballin et al [57], 2020, Sweden	Two-arm RCT; sample size: 77; mean age 71.0 years; 50% female participants; mean BMI 29.2 kg/m^2	• Web-based exercise (n=38) • Supervised exercise (n=39)	Digital tool: Healthy Ageing Initiative; a 10-week intervention consisting of a web-based progressive interval training program (10 weekly videos, 3 sessions per week), while the supervised exercise group participated in in-person sessions.	Significant decreases in BF% and FM in both groups, and in BMI only in the supervised exercise group; no significant between-group differences, except for FM (with a greater reduction in the supervised exercise group).	—
Hyun [58], 2021, Korea	Two-arm RCT; sample size: 16; mean age 38.4 years; 100% female participants; mean BMI 25.6 kg/m^2	• Moderate intensity training (n=8) • High intensity training (n=8)	Digital tool: Real-time video web program; a 8-week real-time video web program based on international PA and exercise guidelines, including bidirectional communication along with instructor feedback. Training sessions lasted 30 minutes (for moderate intensity group) or 50 minutes (for high intensity group), held 3 times weekly, with exercise intensity increasing biweekly for the high intensity group.	Significant decreases in BW, WC, FM, BMI, and BF% in both groups, with significant between-group differences for BMI, WC, and FM (greater reductions in the high intensity group), but not for BW or BF%.	Significant left-hand grip strength improvement in the high intensity group and right-hand in both groups; significant between-group differences.
Seo et al [59], 2023, Korea	Three-arm RCT; sample size: 75; mean age 48.3 years; 100% female participants; mean BMI 25.5 kg/m^2	• Virtual reality exercise (n=25) • Indoor bicycle exercise (n=25) • Control (n=25)	Digital tool: VRFit app; an 8-week virtual reality physical exercise program using an Internet of Things sensor attached to an indoor bicycle, connected to a smartphone, and paired with a head-mounted display for immersive virtual reality exercise. Exercise sessions were conducted 3-5 times weekly, following international guidelines.	Significant decreases in BMI in the virtual reality group, with significant between-group differences (no BMI change in the 2 other groups).	Significant improvement in exercise fun in the virtual reality group, along with significant between-group differences.

[a]PA: physical activity.
[b]RCT: randomized controlled trial.
[c]BW: body weight.
[d]BF%: body fat percentage.
[e]FM: fat mass.
[f]LPA: light physical activity.
[g]MVPA: moderate to vigorous physical activity.
[h]mHealth: mobile health.
[i]Not available.
[j]WC: waist circumference.
[k]WHR: waist-to-hip ratio.
[l]MPA: moderate physical activity.
[m]VPA: vigorous physical activity.

Qualitative Data Synthesis

The duration of the interventions ranged from 8 weeks to 24 months, with an average of 6.6 (SD 5.5) months. Although all studies included interventions involving PA or exercise through action planning, coaching, or goal setting, there were substantial variations in the PA-related methods and techniques used, as well as in the inclusion of other lifestyle components alongside PA. A total of 14 studies [18,35-38,41,44,50,52-54,56,57,59] used digital interventions that delivered structured physical exercise programs, while other studies included flexible PA or exercise routines based on international guidelines. Regarding the type of physical exercise, all studies incorporated aerobic or moderate to vigorous PA (MVPA) as a core component, while 13 studies [18,35,36,40-42,44,50,52-54,56,57] additionally included resistance or strength training. All but 3 studies [44,56,59] incorporated multiple lifestyle components, including diet, sleep, and stress management, with diet being the most commonly used component alongside PA. Dietary guidance was provided in the form of dietary plans, goal setting based on dietary guidelines, nutritional advice, dietary coaching, and behavior change strategies related to eating habits.

Various digital tools, including mobile apps, websites, wearable devices, telecommunication platforms (eg, videoconferencing), virtual reality, or a combination of 2 or more in most studies, were used in the interventions. These tools supported features such as program delivery, coaching, feedback provision, goal

XSL•FO
RenderX

setting, tracking, and progress monitoring. They were implemented in various ways, including automated or manual feedback, self-guided or professionally guided programs, real-time or asynchronous coaching, gamification elements, integration with other health platforms, and different levels of interactivity. This resulted in a diverse range of intervention designs, delivery modes, and user engagement levels.

While 4 studies [35,36,41,47] used a single-arm design, the control groups in the other studies exhibited considerable variability, including active controls (such as alternative digital or nondigital interventions or usual care) [33,34,37-40,42,43,45,46,49,51-58], passive control (such as no intervention or waitlist group) [20,44,48], or a combination of both [16,18,50,59], reflecting a diversity in methodological designs and approaches. All single-arm design studies demonstrated a significant reduction in outcome variables. Similarly, interventional studies comparing digital interventions to passive controls showed favorable effects. However, in studies comparing experimental groups to nondigital active controls, the results were inconsistent, with some reporting greater reductions in outcome variables following digital interventions [37,42,43,45,49,51,59], while others found no significant difference [16,18,38-40,46,50], and one [34] even reported less reduction following digital interventions. Common nondigital interventions included educational booklets, self-care advice, and in-person physical exercise programs.

Among the studies included in this review, 25 evaluated BW as an outcome variable. Overall, 22 studies (with an average intervention duration of 6.5, SD 4.9 months, ranging from 8 weeks to 24 months) reported significant BW reduction (range −1.3 to −8.4 kg) following digital intervention programs [18,20,33,35-43,45-51,53,55,58], while 3 studies found no significant change in BW following a 12-week multicomponent telerehabilitation program [52], a 12-week smartphone mirroring-based telepresence exercise program [54], or a 24 months of mobile health (mHealth) intervention [34]. A total of 10 studies (with an average intervention duration of 9.0, SD 8.4 months, ranging from 3 to 24 months) found significant differences in postintervention BW between their study groups [33,34,37,42,43,45,48-50,55], most of which used active control groups, with one study [48] using passive controls and another [50] using a combination of both. However, 10 studies (with an average intervention duration of 4.7, SD 1.7 months, ranging from 8 weeks to 6 months) reported no significant between-group differences in BW [18,20,38-40,46,52-54,58], most of which used active control groups, with one study [20] using passive controls and another [18] using a combination of both. No clear pattern was observed in BW outcomes when comparing experimental groups to nondigital active controls, with some studies reporting greater BW reduction following digital interventions [37,42,43,45,49], while others found no significant difference [18,38-40,46] or reported less BW reduction following digital interventions [34].

In this review, 23 studies examined BMI as an outcome measure. In total, 19 studies (with an average intervention duration of 5.9, SD 5.2 months, ranging from 8 weeks to 24 months) observed a significant reduction in BMI (range −0.4 to −3.4 kg/m^2) following digital intervention programs [16,20,33,36,37,39,41,42,45-51,53,55,58,59], while 4 studies (with an average intervention duration of 5.1, SD 4.6 months, ranging from 10 weeks to 12 months) reported no significant BMI change following digital interventions [35,52,56,57]. A total of 8 studies (with an average intervention duration of 4.6, SD 3.4 months, ranging from 8 weeks to 12 months) reported significant differences in postintervention BMI between study groups [37,42,45,48-50,58,59], most of which used active control groups, while one study [55] used passive controls and 2 others [50,59] used a combination of both. However, 9 studies (with an average intervention duration of 6.3, SD 6.8 months, ranging from 10 weeks to 24 months) found no significant between-group differences in BMI changes [16,20,33,39,46,52,55-57], most of which used active control groups, with one study [20] using passive controls and another [16] using a combination of both. No clear trend was observed in BMI outcomes when comparing experimental groups to nondigital active controls, with some studies reporting greater BMI reduction following digital interventions [37,42,45,49,59], while others found no significant difference [16,39,46].

A total of 16 studies examined WC as an outcome measure. Overall, 15 studies (with an average intervention duration of 5.7, SD 3.0 months, ranging from 8 weeks to 12 months) found a significant reduction in WC (range −2.1 to −9.2 cm) following digital intervention programs [18,20,35,37-39,41,42,44,46,48,50,51,56,58], while one study [49] showed no significant change following a 3-month multicomponent lifestyle intervention delivered via a mobile app. In total, 8 studies (with an average intervention duration of 5.3, SD 3.1 months, ranging from 8 weeks to 12 months) reported significant between-group differences in postintervention WC [18,37,42,44,48-50,58]; however, 5 studies (with an average intervention duration of 5.4, SD 1.3 months, ranging from 12 weeks to 6 months) indicated no significant between-group differences [20,38,39,46,56]. There was also no clear trend in WC outcomes when comparing experimental groups to nondigital active controls, with 3 studies [37,42,49] reporting greater WC reduction following digital interventions, while 3 others [38,39,46] found no significant difference. WHR was assessed in 6 studies, with 2 studies (conducting 6-month and 3-month interventions) reporting a significant reduction in WHR (−0.02 and −0.01, respectively) [38,50], while the other 4 studies (with an average intervention duration of 6.0, SD 4.2 months, ranging from 12 weeks to 12 months) found no change in WHR following digital intervention programs [36,41,49,52].

In this review, FM was examined as an outcome measure in 7 studies (with an average intervention duration of 7.9, SD 7.8 months, ranging from 8 weeks to 24 months), all of which found a significant reduction (range −0.4 to −6.5 kg) following digital intervention programs [20,33,35,46,52,57,58]. Among them, 3 studies (with an average intervention duration of 3.5, SD 2.2 months, ranging from 8 weeks to 6 months) reported significant between-group differences after the intervention [20,57,58], while 3 others (with an average intervention duration of 11.0, SD 11.3 months, ranging from 12 weeks to 24 months) found no significant difference in FM between their study groups [33,46,52]. In terms of BF%, 9 studies measured this outcome

XSL•FO
RenderX

(with an average intervention duration of 6.8, SD 7.1 months, ranging from 8 weeks to 24 months), all of which found a significant reduction (range −0.3% to −4.1%) from pre- to posttests in digital intervention groups [20,33,36,46,52,54,56-58]. However, the majority (8 studies) reported no significant between-group differences in BF% changes, and one study [36] followed a single-arm design. No clear trend was observed in FM and BF% outcomes when comparing digital-based experimental groups to nondigital active controls.

Regarding PA, 8 studies provided information on changes in MVPA following digital interventions, with 4 reporting significant improvements with an average intervention duration of 9.2 (SD 9.9; range 3-24) months [16,33,43,47] and the other 4 observing no change in MVPA with an average intervention duration of 5.2 (SD 1.5) months [18,20,39,53]. A similar trend of inconsistent results was observed across other PA-related variables, including general PA [37,45,48-50], light PA [18,33,53], moderate PA [20,51], vigorous PA [20,51], daily steps [51,55,56], grip strength [41,54,58], and the 6-minute walk test [41,52]. In addition, 5 other variables, including exercise capacity [44], exercise fun [59], exercise self-efficacy [49], sit-to-stand test [41], and gait speed [41], were each assessed in just one study. Inconsistent trends in postintervention changes were observed in these variables, with some showing significant improvements [44,49,59], while others displayed no significant changes [50], and some revealed mixed results within individual studies [41,54].

Meta-Analysis

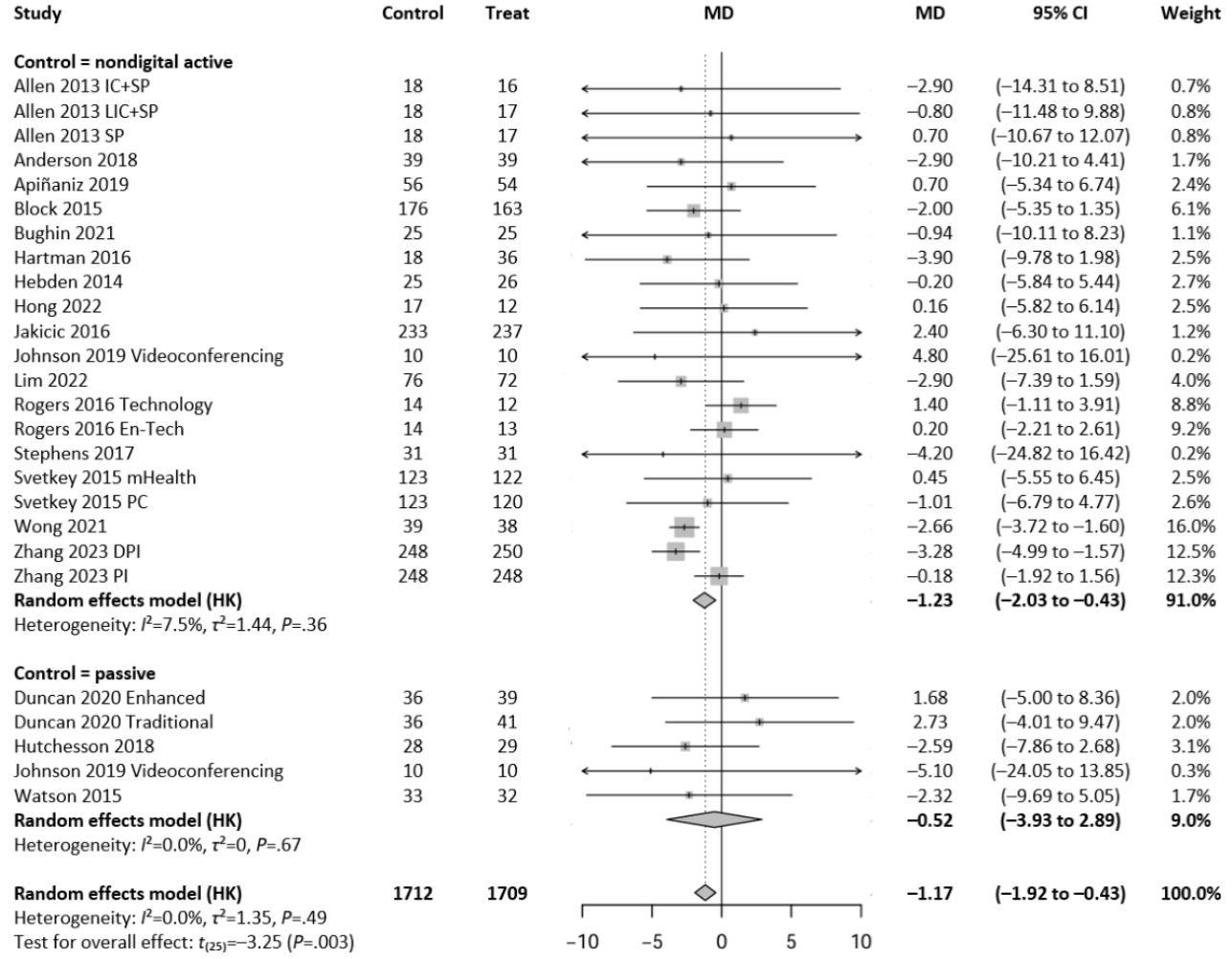
A total of 26 intervention arms from 19 independent studies were included in the meta-analysis assessing the effect of digital exercise interventions on BW. The overall pooled analysis revealed a significant reduction in BW following digital exercise interventions compared to control conditions (MD=−1.17 kg, 95% CI −1.92 to −0.43; P=.003; t_{25}=−3.25). Heterogeneity was not significant (I^2=0.0%; τ^2=1.35; P=.49), indicating consistency across studies. Subgroup analyses based on control type showed a greater effect in studies comparing digital interventions to active (nondigital) controls than to passive controls (MD=−1.23 kg vs −0.52 kg, respectively). Figure 2 [18,20,33,34,37,39,40,42,43,45,46,48-55] shows the results of the meta-analysis assessing the effect of guideline-based digital exercise interventions on BW.

Regarding BMI, a total of 26 intervention arms from 17 independent studies were included in the meta-analysis. A random-effects model revealed a statistically significant reduction in BMI favoring digital interventions (MD=−0.50 kg/m^2, 95% CI −0.82 to −0.19; P=.003; t_{25}=−3.32), with significant heterogeneity across studies (I^2=70.0%; τ^2=0.40; P<.001). Subgroup analyses based on control type showed a greater effect in studies comparing digital interventions to passive controls than to active (nondigital) controls (MD=−0.70 kg/m^2 vs −0.45 kg/m^2, respectively). Figure 3 [16,20,33,37,39,42,45,46,48-53,55,57,59] shows the results of the meta-analysis assessing the effect of guideline-based digital exercise interventions on BMI.

Regarding BF%, a total of 7 intervention arms from 6 independent studies were included in the meta-analysis. The overall analysis revealed no statistically significant difference between interventions and control conditions (MD=−0.08%, 95% CI −0.94 to 0.79; P=.84, t_6=−0.21). Heterogeneity was not significant (I^2=7.4%; τ^2=0.25; P=.37). Figure 4 [20,33,46,52,54,57] shows the results of the meta-analysis assessing the effect of guideline-based digital exercise interventions on BF%.

XSL•FO
RenderX

Figure 2. Forest plot showing the effect of guideline-based digital exercise interventions on body weight across 26 comparisons between digital exercise interventions and control conditions, using a random-effects model with Hartung-Knapp (HK) adjustment. Subgroup analyses are presented for comparisons with nondigital active controls (21 comparisons) and passive controls (5 comparisons), while the last model represents a pooled analysis of all studies (n=1709 intervention participants vs n=1712 controls). Each horizontal line represents a study's 95% CI, with diamond shapes reflecting the pooled estimates. DPI: diet and physical activity intervention; En-Tech: enhanced technology; IC: intensive counseling; LIC: less intensive counseling; MD: mean difference; mHealth: mobile health; PC: personal coaching; PI: physical activity intervention; SP: smartphone.



XSL•FO
RenderX

Figure 3. Forest plot showing the effect of guideline-based digital exercise interventions on BMI across 26 comparisons between interventions and control conditions, using a random-effects model with Hartung-Knapp (HK) adjustment. Subgroup analyses are presented for comparisons with nondigital active controls (20 comparisons) and passive controls (6 comparisons), while the last model represents a pooled analysis of all studies (n=1469 intervention participants vs n=1486 controls). Each horizontal line represents a study's 95% CI, with diamond shapes reflecting the pooled estimates. DPI: diet and physical activity intervention; En-Tech: enhanced technology; IC: intensive counseling; LIC: less intensive counseling; MD: mean difference; PI: physical activity intervention; SP: smartphone only; Virtual-R: virtual reality.

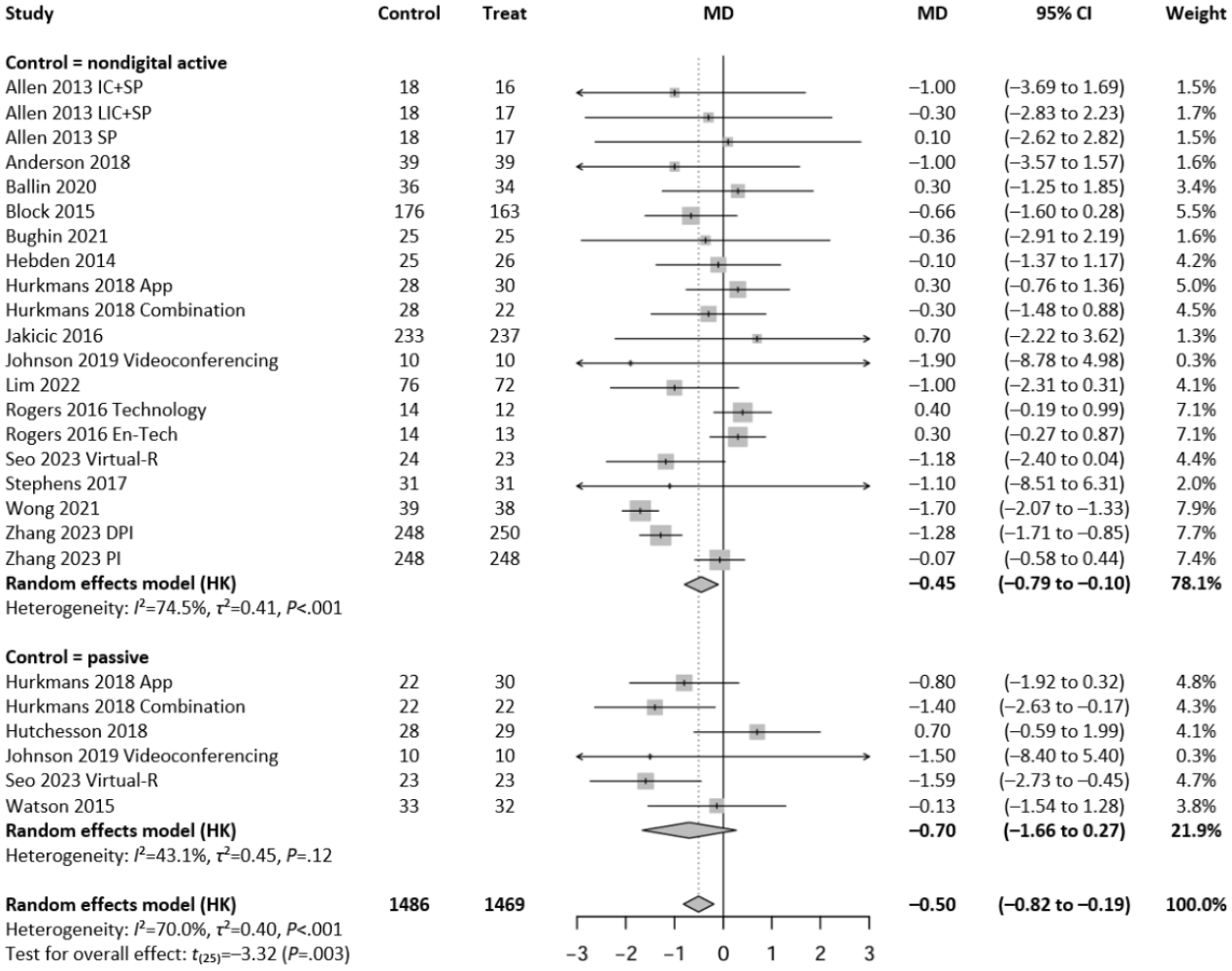
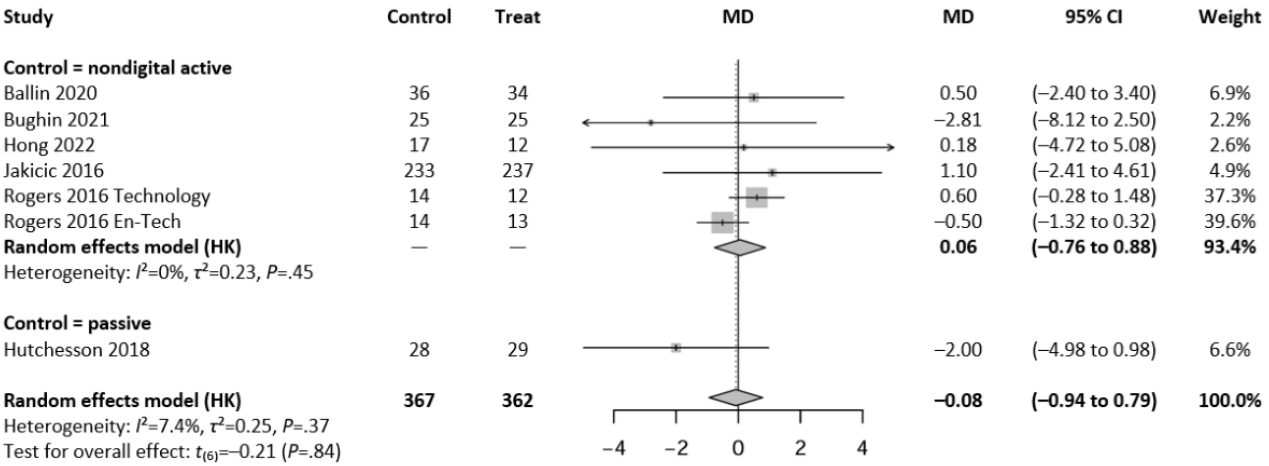


Figure 4. Forest plot showing the effect of guideline-based digital exercise interventions on body fat percentage using a random-effects model with Hartung-Knapp (HK) adjustment. Subgroup analyses are presented for comparisons with nondigital active controls (6 comparisons), while only 1 comparison was available for passive controls. The last model represents a pooled analysis of all studies (n=362 intervention participants vs n=367 controls). Individual study estimates are shown with 95% CIs; diamonds represent pooled estimates. En-Tech: enhanced technology; MD: mean difference.



XSL•FO
RenderX

Risk of Bias

Figure 5 [16,18,20,31-59] summarizes the risk of bias assessment for the included studies, with panel A presenting the overall assessment for RCTs and panel B for nonrandomized studies. Among the 26 RCTs, 4 (15%) were classified as having a "low risk of bias," indicating a strong methodological design with minimal threats to validity. In total, 65% (17/26) of the studies were rated as having "some concerns," suggesting potential sources of bias that could affect the reliability of their findings. In addition, 19% (5/26) of the studies were deemed to have a "high risk of bias," highlighting significant methodological limitations that may compromise the robustness of their results. Among the 4 non-RCTs, 2 (50%) were found to have a "moderate risk of bias," implying methodological shortcomings that could impact their conclusions. The remaining 50% (2/4) of the studies were rated as having "some concerns," indicating the presence of potential biases that warrant caution when interpreting their findings. Excluding studies with a high or serious risk of bias did not substantially alter the overall pattern of results. Similarly, findings from the 4 low-risk-of-bias studies [42,44,45,52] aligned with the general patterns, and the 7 studies with high or serious risk of bias showed the same pattern [16,35-37,40,55,58].

Figure 5. Summary of risk of bias assessment using (A) the Cochrane Risk of Bias (version 2) tool for randomized controlled trials and (B) the Cochrane Risk of Bias in Nonrandomized Studies of Interventions tool for nonrandomized studies.

A

	D1	D2	D3	D4	D5	Overall
Alcántara-Aragón et al (2018)	+	?	?	+	+	?
Allen et al (2013)	+	?	?	+	+	?
Anderson et al (2018)	+	?	?	+	+	?
Apiñaniz et al (2019)	+	?	?	?	+	x
Ballin et al (2020)	+	?	+	+	+	?
Block et al (2015)	+	+	+	+	+	+
Bughin et al (2021)	+	+	+	+	+	+
Duncan et al (2020)	+	?	+	+	?	?
Hartman et al (2016)	+	?	+	+	?	?
Haufe et al (2021)	+	+	+	+	+	+
Hebden et al (2014)	+	?	+	+	?	?
Hong et al (2022)	+	?	?	+	+	?
Hurkmans et al (2018)	?	?	+	+	x	x
Hutchesson et al (2018)	+	+	?	+	+	?
Hyun (2021)	+	?	?	+	?	x
Jakicic et al (2016)	+	?	?	+	+	?
Johnson et al (2019)	?	?	+	?	?	x
Lim et al (2022)	+	+	+	+	+	+
Pressler et al (2010)	+	?	?	+	+	?
Rogers et al (2016)	+	?	?	+	+	?
Seo et al (2023)	+	?	+	?	+	?
Stephens et al (2017)	+	?	+	+	?	?
Svetkey et al (2015)	+	?	+	+	+	?
Watson et al (2015)	?	?	x	+	+	x
Wong et al (2021)	+	?	+	+	?	?
Zhang et al (2023)	+	?	?	+	+	?

D1 (domain 1): Bias arising from the randomization process
D2 (domain 2): Bias due to deviations from intended interventions
D3 (domain 3): Bias due to missing outcome data
D4 (domain 4): Bias in measurement of the outcome
D5 (domain 5): Bias in selection of the reported result

+ Low risk of bias
? Some concerns
x High risk of bias

B

	D1	D2	D3	D4	D5	D6	D7	Overall
Batsis et al (2021)	?	?	+	+	?	?	+	?
Héroux et al (2017)	?	?	+	-	!	+	+	!
Hurst et al (2021)	?	?	+	+	?	+	?	?
Lynch et al (2017)	-	?	?	+	?	?	+	!

D1 (domain 1): Bias due to confounding
D2 (domain 2): Bias in selection of participants into the study
D3 (domain 3): Bias in classification of interventions
D4 (domain 4): Bias due to deviations from intended interventions
D5 (domain 5): Bias due to missing data
D6 (domain 6): Bias in measurement of the outcome
D7 (domain 7): Bias in selection of the reported result

+ Low risk of bias
? Moderate risk of bias
! Serious risk of bias
x Critical risk of bias
- No information

XSL•FO
RenderX

Discussion

Principal Findings

This systematic review and meta-analysis assessed the effectiveness of guideline-based digital exercise interventions in improving body composition in adults with overweight and obesity, aiming to identify gaps in the literature and recommend strategies for developing and implementing future digital health interventions. In addition, alterations in PA were considered as a secondary outcome. The main findings of the meta-analysis were (1) guideline-based digital exercise interventions significantly reduced BW (MD=−1.17 kg; *P*=.003), with greater effects observed when compared to active (nondigital) controls; (2) BMI was also significantly reduced (MD=−0.50 kg/m^2; *P*=.003), with stronger effects observed in studies using passive (waitlist) control groups; and (3) no significant effect was found for BF% (MD=−0.08%; *P*=.84). The main qualitative findings were (1) digital exercise intervention programs led to significant reductions in BW (22/25, 88% studies; range −1.3 to −8.4 kg); BMI (19/23, 83% studies; range −0.4 to −3.4 kg/m^2); WC (15/16, 94% studies; range −2.1 to −9.2 cm); BF% (9/9, 100% studies; range −0.3% to −4.1%); and FM (7/7, 100% studies; range −0.4 to −6.5 kg); (2) inconsistent results were observed for WHR and PA outcomes; (3) no clear trend was identified between intervention duration or type and improvements in outcome variables; and (4) substantial variations were observed in intervention designs, comparators, confounders, and participant characteristics. The digital exercise apps and tools included in this review addressed behavior change through multiple evidence-based strategies grounded in established health behavior theories. Most interventions incorporated goal setting, self-monitoring, and feedback loops, known as core components of behavior change techniques. For instance, studies using apps such as Lose It!, Alive-PD, MetS App, and Balanced App featured personalized goal setting for PA, daily exercise reminders, and tracking tools to enhance user accountability and adherence [18,42,48,49]. Several interventions explicitly applied behavioral theories, such as social cognitive theory, the transtheoretical model, and motivational interviewing, emphasizing constructs such as self-efficacy, stages of readiness, and problem-solving [20,34,45,51]. These constructs were operationalized through coaching, interactive feedback, and gradual PA progression plans. In this context, evidence indicates that human-computer interaction plays a vital role in the effectiveness of digital health interventions, as well-designed interfaces enhance user engagement and adherence through personalized feedback, interactivity, and intuitive navigation [60]. Moreover, some studies used peer interaction or professional coaching via telehealth, which are known to enhance engagement in PA [33,37,47]. Together, these elements show that digital exercise apps, when grounded in PA guidelines and behavior change theory, may extend beyond the delivery of exercise routines to potentially support sustained behavior change by addressing key psychological and social determinants of health behavior.

Body Composition Outcomes

Consistent with this review, evidence from other review studies suggests that mHealth and digital tools can be effective for BW reduction [61-65]. Results from a systematic review [66] indicate that eHealth interventions are generally more effective than no care or usual care interventions, aligning with this meta-analysis, which found a significant average BW loss of approximately 1.17 kg associated with digital exercise interventions. This effect remained significant when compared to active (nondigital) control interventions (MD=−1.23 kg), suggesting that digital platforms may have advantages over traditional, face-to-face programs. These advantages may stem from features such as improved accessibility, personalization, and real-time feedback, all of which can contribute to greater user engagement. Research suggests that the impact of digital interventions on BW reduction may be influenced by factors such as intervention duration and design. In particular, the effectiveness of mHealth interventions for BW reduction tends to increase with longer intervention durations [63,67]. Furthermore, it has been documented that digital interventions incorporating personalized feedback and human interaction are more effective than fully automated programs [66,68]. In this regard, evidence suggests that successful digital interventions often include components such as self-monitoring, social support, goal setting, and prolonged engagement, all of which are crucial for enhancing user engagement and promoting sustained behavior change [69,70].

In line with the results of this review, evidence from comparable review studies suggests that mHealth and digital interventions can be effective in reducing BMI [61-63,71,72], with their impact largely influenced by factors such as the type of intervention, its integration with other health services, and user engagement levels [66,73,74]. In this review, no clear pattern was found between the type and duration of interventions and BMI reductions. However, a stronger effect was observed in studies using passive (waitlist) control groups compared to active controls. A review [63] found that longer intervention duration enhanced the effectiveness of mHealth and digital interventions in reducing BMI. Cultural and regional factors may also indirectly influence the effectiveness of mHealth interventions in BW outcomes; nevertheless, a study evaluating a digital weight loss intervention across multiple countries [75] found similar BW outcomes although levels of participants' engagement varied. These factors may contribute to the high degree of heterogeneity among studies on this topic, as their inclusion adds complexity to the interventions. However, it is important to note that the methodological quality of studies on digital interventions for BW and BMI reduction varies significantly, which can affect the reliability of their findings and the associated review studies [66,76], although excluding studies with low methodological quality did not substantially alter the overall pattern of results in this study.

WC and WHR are 2 important obesity-related measurements used to assess fat distribution. A meta-analysis [61] reported a statistically significant reduction in WC in adults with overweight and obesity following mHealth interventions. Similarly, another review study [72] found a consistent reduction in WC following web- and mobile-based interventions. These findings align with the results of this review, which examines

XSL•FO
RenderX

the effectiveness of guideline-based digital exercise interventions on WC. Results from another systematic review [77] show that mHealth and eHealth interventions, particularly those with personalized feedback, have been associated with reductions in central obesity measures. While there is a lack of similar review studies in the literature assessing the impact of digital interventions on WHR, this review identified 2 studies reporting a reduction in WHR, while 4 other studies reported no change following digital intervention programs. This insignificant trend, which has not been observed in BW, BMI, or WC, may be attributed to the fact that WHR appears to be an inaccurate measure of obesity in individuals with obesity, as both waist and hip circumferences reflect fat accumulation in these individuals [78], and fat loss occurs throughout the body, not in specific regions.

Consistent with the qualitative findings of this systematic review, several studies [63,72] have reported that mHealth and digital interventions can lead to significant reductions in BF%. However, this meta-analysis found no statistically significant difference in the change in BF% between intervention and control groups, with subgroup analysis also revealing no significant effect when compared to active controls. This may be attributed to the limited number of studies assessing BF%, highlighting the need for further research. Consistently, results from a meta-analysis show an insignificant change in BF% following mHealth interventions in adults with overweight and obesity, while FM reduction was significant [61]. The discrepancy in BF% but not in FM, compared to the findings of this review, may be due to the primary focus of this review on the PA component, specifically guideline-based PA, as physical exercise contributes to increased lean mass, thereby having a greater impact on BF%. Research shows that interventions using telehealth and text messaging have also demonstrated positive effects on body fat reduction by providing continuous support and motivation, which can enhance adherence to the lifestyle changes necessary for fat reduction [79]. The use of multiple electronic modalities, such as combining mHealth with other digital tools, tends to yield more significant results in fat reduction compared to single-modality interventions, offering a more comprehensive and engaging intervention experience [80]. Although mHealth and digital interventions demonstrate potential for fat reduction, evidence indicates that the long-term sustainability of these interventions remains uncertain, and their effectiveness may differ based on factors such as socioeconomic status and ethnicity, which could contribute to disparities in health outcomes [81]. Overall, based on the available literature and the present findings, effective digital PA interventions for improving body composition typically incorporate human interaction and sustained engagement, with personalized feedback, social support, and goal setting further enhancing their efficacy.

PA Outcomes

There is limited evidence showing apps to be more effective for PA outcomes than face-to-face interventions [61,82,83]; however, digital tools may offer benefits in terms of cost-effectiveness and the ability to reach a larger and more geographically dispersed population [10,11]. Insufficient and inconsistent results were found for PA outcomes in this review. However, reports from other review studies indicate an increase in general PA [84], daily step counts [65], and weekly MVPA [65] following mHealth interventions. Another review study [85] examining the effects of mHealth interventions on older adults reported an increase in weekly MVPA. A recent umbrella review of mHealth and eHealth interventions [86] reported a significant increase in daily step counts among adult populations. However, 2 systematic reviews found no significant change in general PA [82] or MVPA [61] following mHealth interventions. These discrepancies, which have also been observed in this review, may be explained by variations in intervention design and duration, target populations, and the methods used to measure PA outcomes. A review study [83] highlighted that combining digital media with face-to-face support showed the greatest potential for increasing PA compared to digital-only interventions. This highlights the need for more rigorous research to confirm the long-term benefits of digital interventions in improving PA; however, the methodological quality of the included studies in this review did not alter the results related to PA outcomes.

Limitations and Strengths

This systematic review contains limitations that need to be acknowledged. While the variability in intervention designs and duration has made direct comparisons challenging, variations in participant characteristics (such as baseline BMI, PA levels, sex distributions, age, socioeconomic status, and comorbidities) may have influenced individual study results, and consequently, the findings of this systematic review. In addition, it is important to elaborate on the potential confounding variables that may have influenced the results across the included studies. These confounders, often unevenly distributed or poorly controlled across trials, can influence both the internal and external validity of the findings. For instance, dietary intake, a key determinant of physical and metabolic outcomes, was not consistently reported or controlled in many studies, making it difficult to isolate the effects of the mHealth interventions alone. Similarly, factors such as motivation, digital literacy, and health literacy can substantially influence both adherence and outcomes but were rarely measured or adjusted for. In some populations, cultural factors may also affect attitudes toward PA and technology use, further complicating comparisons across studies conducted in different geographic or cultural contexts. These unmeasured or uncontrolled confounding variables introduce heterogeneity and potential bias into the pooled results, and their presence highlights the need for better standardization and reporting in future mHealth intervention studies. The quality of evidence supporting mHealth interventions should also be considered, as it may vary widely, although this was not observed in this study. Previous systematic reviews and meta-analyses [87,88] have rated the methodological quality of mHealth studies as low to moderate, with only a few achieving high-quality ratings, primarily due to differences in study design, sample size, and outcome measures.

Despite the limitations, this systematic review presents a comprehensive evaluation of guideline-based digital exercise interventions across multiple anthropometric and body composition outcomes, as well as PA-related variables. The inclusion of a variety of intervention formats allows for a broad

XSL•FO
RenderX

understanding of the effectiveness of mHealth and digital approaches. In addition, the review follows a rigorous methodology, ensuring high-quality data analysis and synthesis, thereby contributing to the growing body of evidence-based literature on digital health interventions for improving body composition and promoting PA.

Overall, this review provides insights for individuals, trainers, coaches, health care providers, health organizations, app developers, and policy makers by identifying the potential benefits of apps and digital tools focusing on PA for BW management, thereby contributing to the broader effort to manage overweight and obesity effectively. In addition, the findings highlight areas for future research to optimize the design and implementation of fitness apps for better health outcomes. Future research should focus on the role of personalized and adaptive digital programs, as well as their integration with other lifestyle interventions. By understanding the essential characteristics that enhance the effectiveness of digital health interventions, stakeholders can make informed decisions about integrating them into weight management programs and public health strategies. Meanwhile, clinicians and health care providers can leverage digital technologies to improve accessibility, enhance patient engagement, optimize treatment adherence, and support long-term success in weight management programs.

Conclusions

This study highlights the potential of guideline-based digital exercise interventions for addressing overweight and obesity. While significant reductions in BW and BMI were observed, results were less clear regarding body composition changes. Although longer-duration interventions showed a trend toward greater BW reduction, no clear pattern emerged between the type or duration of interventions and improvements in body composition. Key characteristics that are often identified as critical for successful digital PA interventions include personalized content, regular user engagement, and integration with other health strategies, such as dietary management and social support. While this review highlights the potential benefits of guideline-based digital exercise interventions for weight management, it also provides valuable insights to inform decisions about integrating them into health programs and public health strategies. Overall, digital health interventions can reach a larger population at a lower cost and have been shown to be effective for weight loss; however, evidence of their superiority over traditional methods is limited and inconsistent, suggesting that digital interventions may serve as a complementary approach rather than a definitive replacement for traditional methods. As technology continues to evolve, future interventions have the potential to examine and offer more tailored, engaging, and effective solutions for weight management.

Acknowledgments

There are no professional relationships with companies or manufacturers that benefit from this paper.

This study is funded by a PhD scholarship awarded to MM by the University of Innsbruck, Vice Rectorate for Research (Doktoratsstipendium aus der Nachwuchsförderung 2021 der Universität Innsbruck). The funding agency had no influence on the study design, conduct, data collection and analysis, or the presentation and publication of the findings.

Authors' Contributions

MM contributed to conceptualization, investigation, formal analysis, funding acquisition, methodology, writing—original draft, and writing—review and editing. CD contributed to conceptualization, supervision, investigation (third reviewer), formal analysis, methodology, and writing—review and editing. DT contributed to investigation (second reviewer), formal analysis, methodology, and writing—review and editing. GR contributed to investigation and writing—review and editing. WK contributed to investigation and writing—review and editing. MS contributed to writing—review and editing. TR contributed to general support and writing—review and editing. KW contributed to conceptualization, supervision, investigation, funding acquisition, methodology, and writing—review and editing. All authors critically reviewed and agreed to the final version of the paper to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Keywords used in the database search with Boolean operators: AND (between categories) and OR (between keywords within each category).
[DOCX File , 16 KB - ijmr_v14i1e73656_app1.docx]

Multimedia Appendix 2
PRISMA 2020 checklist.
[DOCX File , 32 KB - ijmr_v14i1e73656_app2.docx]

References

XSL•FO
RenderX

1. GBD 2015 Obesity Collaborators, Afshin A, Forouzanfar MH, Reitsma MB, Sur P, Estep K, et al. Health effects of overweight and obesity in 195 countries over 25 years. N Engl J Med 2017 Jul 06;377(1):13-27 [FREE Full text] [doi: 10.1056/NEJMoa1614362] [Medline: 28604169]
2. Obesity and overweight. World Health Organization. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight [accessed 2025-05-29]
3. Obesity. World Health Organization. URL: https://www.who.int/news-room/facts-in-pictures/detail/6-facts-on-obesity [accessed 2025-05-29]
4. Wiklund P. The role of physical activity and exercise in obesity and weight management: time for critical appraisal. J Sport Health Sci 2016 Jun;5(2):151-154 [FREE Full text] [doi: 10.1016/j.jshs.2016.04.001] [Medline: 30356545]
5. WHO guidelines on physical activity and sedentary behaviour. World Health Organization. URL: https://www.who.int/publications/i/item/9789240015128 [accessed 2025-05-29]
6. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec 25;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
7. Baillot A, Chenail S, Barros Polita N, Simoneau M, Libourel M, Nazon E, et al. Physical activity motives, barriers, and preferences in people with obesity: a systematic review. PLoS One 2021 Jun 23;16(6):e0253114 [FREE Full text] [doi: 10.1371/journal.pone.0253114] [Medline: 34161372]
8. Hoehe MR, Thibaut F. Going digital: how technology use may influence human brains and behavior . Dialogues Clin Neurosci 2022 Apr 01;22(2):93-97. [doi: 10.31887/dcns.2020.22.2/mhoehe]
9. App. Cambridge Dictionary. URL: https://dictionary.cambridge.org/dictionary/english/app [accessed 2025-05-29]
10. Chatterjee A, Prinz A, Gerdes M, Martinez S. Digital interventions on healthy lifestyle management: systematic review. J Med Internet Res 2021 Nov 17;23(11):e26931 [FREE Full text] [doi: 10.2196/26931] [Medline: 34787575]
11. Cai J, Li G. Exercise or lie down? The impact of fitness app use on users' wellbeing. Front Public Health 2023 Jan 10;11:1281323 [FREE Full text] [doi: 10.3389/fpubh.2023.1281323] [Medline: 38269388]
12. Laing BY, Mangione CM, Tseng C, Leng M, Vaisberg E, Mahida M, et al. Effectiveness of a smartphone application for weight loss compared with usual care in overweight primary care patients: a randomized, controlled trial. Ann Intern Med 2014 Nov 18;161(10 Suppl):S5-12 [FREE Full text] [doi: 10.7326/M13-3005] [Medline: 25402403]
13. Balk-Møller NC, Poulsen SK, Larsen TM. Effect of a nine-month web- and app-based workplace intervention to promote healthy lifestyle and weight loss for employees in the social welfare and health care sector: a randomized controlled trial. J Med Internet Res 2017 Apr 10;19(4):e108 [FREE Full text] [doi: 10.2196/jmir.6196] [Medline: 28396303]
14. Martínez-Rodríguez A, Martínez-Olcina M, Mora J, Navarro P, Caturla N, Jones J. New app-based dietary and lifestyle intervention on weight loss and cardiovascular health. Sensors (Basel) 2022 Jan 20;22(3):768 [FREE Full text] [doi: 10.3390/s22030768] [Medline: 35161515]
15. Cho SM, Lee JH, Shim JS, Yeom H, Lee SJ, Jeon YW, et al. Effect of smartphone-based lifestyle coaching app on community-dwelling population with moderate metabolic abnormalities: randomized controlled trial. J Med Internet Res 2020 Oct 09;22(10):e17435 [FREE Full text] [doi: 10.2196/17435] [Medline: 33034564]
16. Hurkmans E, Matthys C, Bogaerts A, Scheys L, Devloo K, Seghers J. Face-to-face versus mobile versus blended weight loss program: randomized clinical trial. JMIR Mhealth Uhealth 2018 Jan 11;6(1):e14 [FREE Full text] [doi: 10.2196/mhealth.7713] [Medline: 29326093]
17. Nakata Y, Sasai H, Gosho M, Kobayashi H, Shi Y, Ohigashi T, et al. A smartphone healthcare application, CALO mama Plus, to promote weight loss: a randomized controlled trial. Nutrients 2022 Nov 02;14(21):4608 [FREE Full text] [doi: 10.3390/nu14214608] [Medline: 36364870]
18. Duncan MJ, Fenton S, Brown WJ, Collins CE, Glozier N, Kolt GS, et al. Efficacy of a multi-component m-Health weight-loss intervention in overweight and obese adults: a randomised controlled trial. Int J Environ Res Public Health 2020 Aug 26;17(17):6200 [FREE Full text] [doi: 10.3390/ijerph17176200] [Medline: 32859100]
19. Hernández-Reyes A, Cámara-Martos F, Molina Recio G, Molina-Luque R, Romero-Saldaña M, Moreno Rojas R. Push notifications from a mobile app to improve the body composition of overweight or obese women: randomized controlled trial. JMIR Mhealth Uhealth 2020 Feb 12;8(2):e13747 [FREE Full text] [doi: 10.2196/13747] [Medline: 32049065]
20. Hutchesson MJ, Callister R, Morgan PJ, Pranata I, Clarke ED, Skinner G, et al. A targeted and tailored eHealth weight loss program for young women: the be positive be health randomized controlled trial. Healthcare (Basel) 2018 May 02;6(2):39 [FREE Full text] [doi: 10.3390/healthcare6020039] [Medline: 29724054]
21. Lugones-Sanchez C, Sanchez-Calavera MA, Repiso-Gento I, Adalia EG, Ramirez-Manent JI, Agudo-Conde C, EVIDENT 3 Investigators. Effectiveness of an mHealth intervention combining a smartphone app and smart band on body composition in an overweight and obese population: randomized controlled trial (EVIDENT 3 study). JMIR Mhealth Uhealth 2020 Nov 26;8(11):e21771 [FREE Full text] [doi: 10.2196/21771] [Medline: 33242020]
22. Martin CK, Miller AC, Thomas DM, Champagne CM, Han H, Church T. Efficacy of SmartLoss, a smartphone-based weight loss intervention: results from a randomized controlled trial. Obesity (Silver Spring) 2015 May 28;23(5):935-942 [FREE Full text] [doi: 10.1002/oby.21063] [Medline: 25919921]

XSL•FO
RenderX

23. 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]
24. 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 05;37(6):387. [doi: 10.1136/emermed-2020-209567] [Medline: 32253195]
25. Body mass index (BMI). World Health Organization. URL: https://www.who.int/data/gho/data/themes/topics/topic-details/GHO/body-mass-index [accessed 2025-05-29]
26. The Asia-Pacific perspective: redefining obesity and its treatment. World Health Organization. 2000. URL: https://iris.who.int/handle/10665/206936 [accessed 2025-05-29]
27. Potter AW, Chin GC, Looney DP, Friedl KE. Defining overweight and obesity by percent body fat instead of body mass index. J Clin Endocrinol Metab 2025 Mar 17;110(4):e1103-e1107. [doi: 10.1210/clinem/dgae341] [Medline: 38747476]
28. Waist circumference and waist-hip ratio: report of a WHO expert consultation. World Health Organization. 2011. URL: https://www.who.int/publications/i/item/9789241501491 [accessed 2025-05-29]
29. 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 [FREE Full text] [Medline: 3920711]
30. Dasso NA. How is exercise different from physical activity? A concept analysis. Nurs Forum 2019 Jan;54(1):45-52. [doi: 10.1111/nuf.12296] [Medline: 30332516]
31. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
32. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. BMJ 2016 Oct 12;355:i4919 [FREE Full text] [doi: 10.1136/bmj.i4919] [Medline: 27733354]
33. Jakicic JM, Davis KK, Rogers RJ, King WC, Marcus MD, Helsel D, et al. Effect of wearable technology combined with a lifestyle intervention on long-term weight loss: the IDEA randomized clinical trial. JAMA 2016 Sep 20;316(11):1161-1171 [FREE Full text] [doi: 10.1001/jama.2016.12858] [Medline: 27654602]
34. Svetkey LP, Batch BC, Lin PH, Intille SS, Corsino L, Tyson CC, et al. Cell phone intervention for you (CITY): a randomized, controlled trial of behavioral weight loss intervention for young adults using mobile technology. Obesity (Silver Spring) 2015 Nov;23(11):2133-2141 [FREE Full text] [doi: 10.1002/oby.21226] [Medline: 26530929]
35. Héroux M, Watt M, McGuire KQ, Berardi JM. A personalized, multi-platform nutrition, exercise, and lifestyle coaching program: a pilot in women. Internet Interv 2017 Mar;7:16-22 [FREE Full text] [doi: 10.1016/j.invent.2016.12.002] [Medline: 30135822]
36. Lynch SM, Stricker CT, Brown JC, Berardi JM, Vaughn D, Domchek S, et al. Evaluation of a web-based weight loss intervention in overweight cancer survivors aged 50 years and younger. Obes Sci Pract 2017 Mar 24;3(1):83-94 [FREE Full text] [doi: 10.1002/osp4.98] [Medline: 28392934]
37. Watson S, Woodside JV, Ware LJ, Hunter SJ, McGrath A, Cardwell CR, et al. Effect of a web-based behavior change program on weight loss and cardiovascular risk factors in overweight and obese adults at high risk of developing cardiovascular disease: randomized controlled trial. J Med Internet Res 2015 Jul 16;17(7):e177 [FREE Full text] [doi: 10.2196/jmir.3828] [Medline: 26183659]
38. Alcántara-Aragón V, Rodrigo-Cano S, Lupianez-Barbero A, Martinez MJ, Martinez C, Tapia J, et al. Web support for weight-loss interventions: PREDIRCAM2 clinical trial baseline characteristics and preliminary results. Diabetes Technol Ther 2018 May;20(5):380-385. [doi: 10.1089/dia.2017.0456] [Medline: 29620918]
39. Allen JK, Stephens J, Dennison Himmelfarb CR, Stewart KJ, Hauck S. Randomized controlled pilot study testing use of smartphone technology for obesity treatment. J Obes 2013;2013:151597 [FREE Full text] [doi: 10.1155/2013/151597] [Medline: 24392223]
40. Apiñaniz A, Cobos-Campos R, Sáez de Lafuente-Moríñigo A, Parraza N, Aizpuru F, Pérez I, et al. Effectiveness of randomized controlled trial of a mobile app to promote healthy lifestyle in obese and overweight patients. Fam Pract 2019 Nov 18;36(6):699-705. [doi: 10.1093/fampra/cmz020] [Medline: 31093681]
41. Batsis JA, Petersen CL, Clark MM, Cook SB, Kotz D, Gooding TL, et al. Feasibility and acceptability of a technology-based, rural weight management intervention in older adults with obesity. BMC Geriatr 2021 Jan 12;21(1):44 [FREE Full text] [doi: 10.1186/s12877-020-01978-x] [Medline: 33435877]
42. Block G, Azar KM, Romanelli RJ, Block TJ, Hopkins D, Carpenter HA, et al. Diabetes prevention and weight loss with a fully automated behavioral intervention by email, web, and mobile phone: a randomized controlled trial among persons with prediabetes. J Med Internet Res 2015 Oct 23;17(10):e240 [FREE Full text] [doi: 10.2196/jmir.4897] [Medline: 26499966]
43. Hartman SJ, Nelson SH, Cadmus-Bertram LA, Patterson RE, Parker BA, Pierce JP. Technology- and phone-based weight loss intervention: pilot RCT in women at elevated breast cancer risk. Am J Prev Med 2016 Nov;51(5):714-721 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.024] [Medline: 27593420]

XSL•FO
RenderX

44. Haufe S, Hupa-Breier KL, Bayerle P, Boeck HT, Rolff S, Sundermeier T, et al. Telemonitoring-supported exercise training in employees with metabolic syndrome improves liver inflammation and fibrosis. Clin Transl Gastroenterol 2021 Jun 18;12(6):e00371 [FREE Full text] [doi: 10.14309/ctg.0000000000000371] [Medline: 34140456]
45. Lim SL, Ong KW, Johal J, Han CY, Yap QV, Chan YH, et al. A smartphone app-based lifestyle change program for prediabetes (D'LITE Study) in a multiethnic Asian population: a randomized controlled trial. Front Nutr 2022 Jan 24;8:780567 [FREE Full text] [doi: 10.3389/fnut.2021.780567] [Medline: 35141265]
46. Rogers RJ, Lang W, Barone Gibbs B, Davis KK, Burke LE, Kovacs SJ, et al. Applying a technology-based system for weight loss in adults with obesity. Obes Sci Pract 2016 Mar 23;2(1):3-12 [FREE Full text] [doi: 10.1002/osp4.18] [Medline: 27812375]
47. Hurst I, Bixenstine PJ, Casillas C, Rasmussen A, Grossman S, Le M, et al. A telehealth intervention to promote weight loss and physical activity in overweight primary care patients. Healthc (Amst) 2021 Mar;9(1):100509. [doi: 10.1016/j.hjdsi.2020.100509] [Medline: 33517179]
48. Stephens JD, Yager AM, Allen J. Smartphone technology and text messaging for weight loss in young adults: a randomized controlled trial. J Cardiovasc Nurs 2017;32(1):39-46 [FREE Full text] [doi: 10.1097/JCN.0000000000000307] [Medline: 26646593]
49. Wong EM, Leung DY, Tam HL, Wang Q, Yeung KW, Leung AY. The effect of a lifestyle intervention program using a mobile application for adults with metabolic syndrome, versus the effect of a program using a booklet: a pilot randomized controlled trial. Clin Interv Aging 2021;16:633-644 [FREE Full text] [doi: 10.2147/CIA.S303920] [Medline: 33888981]
50. Zhang N, Zhou M, Li M, Ma G. Effects of smartphone-based remote interventions on dietary intake, physical activity, weight control, and related health benefits among the older population with overweight and obesity in China: randomized controlled trial. J Med Internet Res 2023 Apr 28;25:e41926 [FREE Full text] [doi: 10.2196/41926] [Medline: 37115608]
51. Anderson AS, Dunlop J, Gallant S, Macleod M, Miedzybrodzka Z, Mutrie N, et al. Feasibility study to assess the impact of a lifestyle intervention ('LivingWELL') in people having an assessment of their family history of colorectal or breast cancer. BMJ Open 2018 Feb 01;8(2):e019410 [FREE Full text] [doi: 10.1136/bmjopen-2017-019410] [Medline: 29391383]
52. Bughin F, Bui G, Ayoub B, Blervaque L, Saey D, Avignon A, et al. Impact of a mobile telerehabilitation solution on metabolic health outcomes and rehabilitation adherence in patients with obesity: randomized controlled trial. JMIR Mhealth Uhealth 2021 Dec 06;9(12):e28242 [FREE Full text] [doi: 10.2196/28242] [Medline: 34874887]
53. Hebden L, Cook A, van der Ploeg HP, King L, Bauman A, Allman-Farinelli M. A mobile health intervention for weight management among young adults: a pilot randomised controlled trial. J Hum Nutr Diet 2014 Aug 29;27(4):322-332. [doi: 10.1111/jhn.12155] [Medline: 23992038]
54. Hong J, Kim SW, Joo H, Kong H. Effects of smartphone mirroring-based telepresence exercise on body composition and physical function in obese older women. Aging Clin Exp Res 2022 May;34(5):1113-1121 [FREE Full text] [doi: 10.1007/s40520-021-02033-4] [Medline: 35028918]
55. Johnson KE, Alencar MK, Coakley KE, Swift DL, Cole NH, Mermier CM, et al. Telemedicine-based health coaching is effective for inducing weight loss and improving metabolic markers. Telemed J E Health 2019 Feb;25(2):85-92 [FREE Full text] [doi: 10.1089/tmj.2018.0002] [Medline: 29847222]
56. Pressler A, Knebel U, Esch S, Kölbl D, Esefeld K, Scherr J, et al. An internet-delivered exercise intervention for workplace health promotion in overweight sedentary employees: a randomized trial. Prev Med 2010 Sep;51(3-4):234-239. [doi: 10.1016/j.ypmed.2010.07.008] [Medline: 20638409]
57. Ballin M, Hult A, Björk S, Lundberg E, Nordström P, Nordström A. Web-based exercise versus supervised exercise for decreasing visceral adipose tissue in older adults with central obesity: a randomized controlled trial. BMC Geriatr 2020 May 12;20(1):173 [FREE Full text] [doi: 10.1186/s12877-020-01577-w] [Medline: 32398024]
58. Hyun AH. Effect of real-time online high-intensity interval training on physiological and physical parameters for abdominally obese women: a randomized pilot study. Appl Sci 2021 Dec 20;11(24):12129. [doi: 10.3390/app112412129]
59. Seo EY, Kim YS, Lee YJ, Hur MH. Virtual reality exercise program effects on body mass index, depression, exercise fun and exercise immersion in overweight middle-aged women: a randomized controlled trial. Int J Environ Res Public Health 2023 Jan 04;20(2):900 [FREE Full text] [doi: 10.3390/ijerph20020900] [Medline: 36673656]
60. Blandford A. HCI for health and wellbeing: challenges and opportunities. Int J Hum Comput Stud 2019 Nov;131:41-51. [doi: 10.1016/j.ijhcs.2019.06.007]
61. Li S, Zhou Y, Ma H, Wang A, Tang X, Pei R, et al. Effects of mobile applications on weight-related, behavior and metabolic outcomes for adults with overweight and obesity: a systematic review and meta-analysis. Stud Health Technol Inform 2024 Jul 24;315:386-391. [doi: 10.3233/SHTI240175] [Medline: 39049288]
62. Liu F, Kong X, Cao J, Chen S, Li C, Huang J, et al. Mobile phone intervention and weight loss among overweight and obese adults: a meta-analysis of randomized controlled trials. Am J Epidemiol 2015 Mar 01;181(5):337-348 [FREE Full text] [doi: 10.1093/aje/kwu260] [Medline: 25673817]
63. Yen HY, Jin G, Chiu HL. Smartphone app-based interventions targeting physical activity for weight management: a meta-analysis of randomized controlled trials. Int J Nurs Stud 2023 Jan;137:104384. [doi: 10.1016/j.ijnurstu.2022.104384] [Medline: 36379104]

XSL•FO
RenderX

64. Neve M, Morgan PJ, Jones PR, Collins CE. Effectiveness of web-based interventions in achieving weight loss and weight loss maintenance in overweight and obese adults: a systematic review with meta-analysis. Obes Rev 2010 Apr;11(4):306-321. [doi: 10.1111/j.1467-789X.2009.00646.x] [Medline: 19754633]
65. Dehghan Ghahfarokhi A, Vosadi E, Barzegar H, Saatchian V. The effect of wearable and smartphone applications on physical activity, quality of life, and cardiovascular health outcomes in overweight/obese adults: a systematic review and meta-analysis of randomized controlled trials. Biol Res Nurs 2022 Oct;24(4):503-518. [doi: 10.1177/10998004221099556] [Medline: 35535558]
66. Kupila SK, Joki A, Suojanen LU, Pietiläinen KH. The effectiveness of eHealth interventions for weight loss and weight loss maintenance in adults with overweight or obesity: a systematic review of systematic reviews. Curr Obes Rep 2023 Sep 24;12(3):371-394 [FREE Full text] [doi: 10.1007/s13679-023-00515-2] [Medline: 37354334]
67. Aisyiyah FN, Murti B, Pamungkasari EP. Meta-analysis: the effectiveness of mHealth mobile application use to promote physical activity and ideal body weight in adult with overweight. J Health Promot Behav 2022;7(3):208-224. [doi: 10.26911/thejhpb.2022.07.03.04]
68. Qin Y, Wang X, Namkoong K. A meta-analysis of the overall effect of mHealth physical activity interventions for weight loss and the moderating effect of behavioral change theories, techniques, and mobile technologies. Mob Media Commun 2021 Nov 16;10(2):337-359. [doi: 10.1177/20501579211054929]
69. Shi Y, Wakaba K, Kiyohara K, Hayashi F, Tsushita K, Nakata Y. Effectiveness and components of web-based interventions on weight changes in adults who were overweight and obese: a systematic review with meta-analyses. Nutrients 2022 Dec 30;15(1):179 [FREE Full text] [doi: 10.3390/nu15010179] [Medline: 36615836]
70. Berry R, Kassavou A, Sutton S. Does self-monitoring diet and physical activity behaviors using digital technology support adults with obesity or overweight to lose weight? A systematic literature review with meta-analysis. Obes Rev 2021 Oct 30;22(10):e13306. [doi: 10.1111/obr.13306] [Medline: 34192411]
71. Beleigoli AM, Andrade AQ, Cançado AG, Paulo MN, Diniz MD, Ribeiro AL. Web-based digital health interventions for weight loss and lifestyle habit changes in overweight and obese adults: systematic review and meta-analysis. J Med Internet Res 2019 Jan 08;21(1):e298 [FREE Full text] [doi: 10.2196/jmir.9609] [Medline: 30622090]
72. Lahtio H, Rintala A, Immonen J, Sjögren T. The effectiveness of physical activity-promoting web- and mobile-based distance weight loss interventions on body composition in rehabilitation settings: systematic review, meta-analysis, and meta-regression analysis. J Med Internet Res 2022 Mar 24;24(3):e25906 [FREE Full text] [doi: 10.2196/25906] [Medline: 35323126]
73. Berry MP, Chwyl C, Metzler AL, Sun JH, Dart H, Forman EM. Associations between behaviour change technique clusters and weight loss outcomes of automated digital interventions: a systematic review and meta-regression. Health Psychol Rev 2023 Dec 06;17(4):521-549. [doi: 10.1080/17437199.2022.2125038] [Medline: 36102170]
74. 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]
75. Yang Q, Mitchell ES, Ho AS, DeLuca L, Behr H, Michaelides A. Cross-national outcomes of a digital weight loss intervention in the United States, Canada, United Kingdom and Ireland, and Australia and New Zealand: a retrospective analysis. Front Public Health 2021 Jun 10;9:604937 [FREE Full text] [doi: 10.3389/fpubh.2021.604937] [Medline: 34178911]
76. Han K, Yu M, Kehinde O. Effectiveness of different online intervention modalities for middle-aged adults with overweight and obesity: a 20-year systematic review and meta-analysis. J Prev (2022) 2024 Feb;45(1):123-157. [doi: 10.1007/s10935-023-00761-z] [Medline: 38114773]
77. O'Boyle J, Davidson P. The effects of mHealth versus eHealth on weight loss in adults: a systematic review. Top Clin Nutr 2022;37(1):69-84. [doi: 10.1097/tin.0000000000000268]
78. Daniel M, Martin AD, Drinkwater DT, Clarys JP, Marfell-Jones MJ. Waist-to-hip ratio and adipose tissue distribution: contribution of subcutaneous adiposity. Am J Hum Biol 2003 Apr 14;15(3):428-432. [doi: 10.1002/ajhb.10165] [Medline: 12704718]
79. Bezerra da Silva LA, De Melo RC, De Araújo BC, Toma TS, Luquine Júnior CD, Milhomens LD, et al. Telehealth and information/communication technologies in the care of overweight and obese adults: a rapid review. Latin Am J Telehealth 2023 Oct 30;9(3):260-283. [doi: 10.32443/2175-2990(2022)431]
80. Lorenzo E, O'Neal AL, Garcia LC, Mendoza K, Lee RE. Electronic health interventions for type 2 diabetes and obesity in Hispanic or Latino adults: a systematic review of English and Spanish studies. Diabetes Spectr 2024;37(1):65-85. [doi: 10.2337/ds22-0083] [Medline: 38385094]
81. Sithole BR, Pappas Y, Randhawa G. eHealth in obesity care. Clin Med (Lond) 2023 Jul;23(4):347-352 [FREE Full text] [doi: 10.7861/clinmed.2023-0145] [Medline: 37524418]
82. Islam MM, Poly TN, Walther BA, Jack Li YC. Use of mobile phone app interventions to promote weight loss: meta-analysis. JMIR Mhealth Uhealth 2020 Jul 22;8(7):e17039 [FREE Full text] [doi: 10.2196/17039] [Medline: 32706724]
83. Ziegeldorf A, Wagner P, Wulff H. Effects of media-assisted therapeutic approaches on physical activity of obese adults: a systematic review. BMC Endocr Disord 2020 Feb 21;20(1):28 [FREE Full text] [doi: 10.1186/s12902-020-0505-x] [Medline: 32085762]

XSL•FO
RenderX

84. Kim HN, Seo K. Smartphone-based health program for improving physical activity and tackling obesity for young adults: a systematic review and meta-analysis. Int J Environ Res Public Health 2019 Dec 18;17(1):15 [FREE Full text] [doi: 10.3390/ijerph17010015] [Medline: 31861359]
85. Núñez de Arenas-Arroyo S, Cavero-Redondo I, Alvarez-Bueno C, Sequí-Domínguez I, Reina-Gutiérrez S, Martínez-Vizcaíno V. Effect of eHealth to increase physical activity in healthy adults over 55 years: a systematic review and meta-analysis. Scand J Med Sci Sports 2021 Apr;31(4):776-789. [doi: 10.1111/sms.13903] [Medline: 33280182]
86. Singh B, Ahmed M, Staiano AE, Gough C, Petersen J, Vandelanotte C, et al. A systematic umbrella review and meta-meta-analysis of eHealth and mHealth interventions for improving lifestyle behaviours. NPJ Digit Med 2024 Jul 05;7(1):179 [FREE Full text] [doi: 10.1038/s41746-024-01172-y] [Medline: 38969775]
87. Chong SO, Pedron S, Abdelmalak N, Laxy M, Stephan A. An umbrella review of effectiveness and efficacy trials for app-based health interventions. NPJ Digit Med 2023 Dec 16;6(1):233 [FREE Full text] [doi: 10.1038/s41746-023-00981-x] [Medline: 38104213]
88. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]

Abbreviations

BF%: body fat percentage
BW: body weight
FM: fat mass
MD: mean difference
MeSH: Medical Subject Headings
mHealth: mobile health
MVPA: moderate to vigorous physical activity
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
WC: waist circumference
WHO: World Health Organization
WHR: waist-to-hip ratio

Edited by T de Azevedo Cardoso; submitted 09.03.25; peer-reviewed by C Br, Q Wang; comments to author 25.04.25; revised version received 26.05.25; accepted 27.05.25; published 07.08.25.

Please cite as:
Motevalli M, Drenowatz C, Tanous D, Ruedl G, Kirschner W, Schauer M, Rosemann T, Wirnitzer K
Guideline-Based Digital Exercise Interventions for Reducing Body Weight and Fat and Promoting Physical Activity in Adults With Overweight and Obesity: Systematic Review and Meta-Analysis
Interact J Med Res 2025;14:e73656
URL: https://www.i-jmr.org/2025/1/e73656
doi:10.2196/73656
PMID:40773286

©Mohamad Motevalli, Clemens Drenowatz, Derrick Tanous, Gerhard Ruedl, Werner Kirschner, Markus Schauer, Thomas Rosemann, Katharina Wirnitzer. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 07.08.2025. 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

Cross-Platform Availability of Smartphone Sensors for Depression Indication Systems: Mixed-Methods Umbrella Review

Johannes Leimhofer[1,2], MSc; Milica Petrovic[1], PhD; Andreas Dominik[3], Prof Dr; Dominik Heider[4], Prof Dr; Ulrich Hegerl[1], Prof Dr Med

[1]Research Center of German Foundation for Depression and Suicide Prevention, Department of Psychiatry, Psychosomatics and Psychotherapy, Goethe University Frankfurt, Frankfurt am Main, Germany

[2]Institute of Computer Science, Faculty of Mathematics and Natural Sciences, Heinrich Heine University Düsseldorf, Düsseldorf, Germany

[3]Life Science Informatics Group, Department of Mathematics, Natural Sciences and Informatics, University of Applied Sciences, Giessen, Germany

[4]Institute of Medical Informatics, University of Münster, Münster, Germany

Corresponding Author:
Dominik Heider, Prof Dr
Institute of Medical Informatics
University of Münster
Albert-Schweitzer-Campus 1, Gebäude A11
Münster, 48149
Germany
Phone: 49 251 830
Email: dominik.heider@uni-muenster.de

Abstract

Background: A popular trend in depression forecasting research is the development of machine learning models trained with various types of smartphone sensor data and periodic self-ratings to derive early indications of changes in depression severity. While most works focus on model performance, there is little concern about the universal usability and reliable operation of such systems across smartphone platforms. This review serves as foundational work for the MENTINA clinical trial, which investigates smartphone-based health self-management for depression. The usability and reliability of mobile apps for depression are commonly perceived through the lens of the approaches and interventions offered rather than the reliability of the built-in mobile phone functions to support effortless and exact delivery of intended interventions.

Objective: This work aimed to synthesize existing systematic reviews to identify smartphone sensor modalities used in mental health monitoring and, building on this foundation, assess the cross-platform availability of these data streams using PhoneDB to inform the design and implementation of digital depression indication systems.

Methods: To identify the already used hardware and software sensors and their purposes in mental health monitoring, an umbrella review was conducted. Three electronic databases, including PubMed, Web of Science Core Collection, and Scopus, were searched using smartphone, sensor data, and depression keyword combination to retrieve relevant literature reviews published within the last 5 years (2019-2024). Once the initial search was completed, the extracted hardware sensors were checked for availability on Android and iOS smartphones by analyzing device specifications in PhoneDB over the last 10 years.

Results: The extracted data streams observed across the 9 included studies covered 16 hardware and 3 software data streams. Hardware data streams included accelerometers, barometers, battery levels, Bluetooth, cameras, cellular networks, GPSs, gyroscopes, humidity, light sensors, magnetometers, proximity sensors, sound sensors, step counters, temperature sensors, and Wi-Fi. Software data streams included app usage, call and message logs, and screen status. Hardware component availability on Android and iOS systems showed the changes in component trends from 2014 to 2024 as of September 2024, with the accelerometers, batteries, cameras, and GPSs remaining consistent on Android and iOS, while components such as gyroscopes, step counters, and barometers gradually increased over the years, particularly on Android.

Conclusions: Multiple data streams identified in the literature review showed a consistent increase in availability over time, enabling improved use of these outputs for depression forecasting and the training of machine learning models with diverse smartphone data, including sensor-derived information. For more precise and reliable data to be used in the mental health field, particularly in critical areas such as tracking and predicting changes in depression severity, further research is required to streamline

XSL•FO
RenderX

smartphone data across varying mobile hardware and software configurations to provide reliable output for digital mental health purposes.

(Interact J Med Res 2025;14:e69686) doi:10.2196/69686

KEYWORDS

umbrella review; mobile phone; smartphone; sensor data; digital health; depression; cross-platform; data availability

Introduction

Background

The vast variety of personally generated data can provide useful insights that empower individuals with depression to better manage experienced symptoms. Smartphone apps allow for the easy collection of personalized data over long time periods and enable the use of powerful machine learning models that can potentially identify individual data patterns related to the severity of depression or predict changes in symptoms and the probability of relapse or remission. Periodic self-ratings concerning depression-related symptoms and life events are mostly collected actively via questionnaires, whereas data from hardware and software components are mostly recorded passively to derive biomarkers, behavior patterns (eg, phone and app usage) and external moderating factors (eg, environmental conditions).

Several promising works are already using smartphone data to train supervised machine learning models for mental health monitoring [1-6], but only a few are based on continual data from studies lasting longer than 30 days. The existing studies show insufficient consideration of the reliability and practicability of the chosen algorithms in real-world scenarios. Once models are deployed and operated, the performance of model prediction results may vary depending on the algorithm's capability to handle missing and erroneous data due to reasons such as varying smartphone platforms or data availability. Moreover, the variety of installed hardware and software components and the steadily evolving technology in this field make it even more challenging to build fail-safe applications that work across various environments in the present and future.

Recent literature lists plenty of passively collected smartphone data streams mostly used in a transformed manner for mental health monitoring. While hardware components like accelerometers, ambient light sensors, Bluetooth, cameras, GPSs, gyroscopes, microphones, and Wi-Fi are among the most used smartphone data sources [7-15], hardware components like barometers [8] and temperature [13] sensors are named only a few times in relevant works, but can also play an important role in the context of mental health monitoring. Most sensors are not directly used for depression modeling but rather serve as a proxy for key figures that are correlated with depression symptoms. Typical key figures derived from smartphone hardware components data are estimates on physical activity, sleep duration, sociability, mobility, circadian rhythms, stress, and the environment recognition [7-15]. The use of sensors for deriving key figures varies widely across the literature. For example, sociability has been inferred using microphones, analyzing silence, noise, and voices [14], or by detecting nearby devices via Bluetooth or Wi-Fi [10]. The term "environment" has been applied inconsistently in the literature, encompassing aspects such as an individual's visual and acoustic surroundings as well as environmental parameters like climatic conditions. For instance, microphones have been used to identify acoustic environments to infer stress [15], while temperature, humidity, and accelerometer sensors have been used to assess environmental hazards [13].

Besides hardware components, smartphone operating systems also provide valuable information for mental health monitoring. Software sensors measuring smartphone and app usage offer key figures on sociability, distractedness, stress, and mood estimates in the case of access permissions to app, phone call, and text message logs [7-15].

Given the presence of multiple existing systematic reviews on smartphone sensing for depression, an umbrella review was chosen to synthesize high-level evidence and critically appraise the quality of these sources—an important step for informing evidence-based mental health app development. By identifying consistently reported sensor modalities and methodological patterns, this review provides a foundation for informed decisions related to recruitment and data collection setup in upcoming clinical studies using smartphone-based mental health monitoring. In particular the results of this work can serve as a comprehensive path for the setting up of the MENTINA trial. The subproject MENTINA trial is part of the European Union–funded research project MENTBEST [16] and aims to investigate the value of data-driven interventions in the context of management of depression severity. Participants of the trial will use an app for 1 year that collects data from smartphone sensors passively while requiring the users to periodically give feedback on personal well-being via self-reports. Besides providing biweekly feedback using the Patient Health Questionnaire 9 [17], participants will also be asked to provide daily ratings on their mental health condition by answering the Patient Health Questionnaire 2 [18] throughout the trial. The app will be available on Android and iOS smartphones and is based on the Monsenso app developed by the Danish digital health solution provider [19-21].

Objectives

In alignment with the preparation of the MENTINA trial, the following research questions should be answered in the course of this work:

1. Which data streams provided by smartphones have been used in the literature to build models for depression monitoring?
2. Which smartphone data streams are available across smartphone vendors and operating systems to build reliable platform-independent software systems for depression monitoring?

XSL•FO
RenderX

3. Which proxies have been constructed from raw smartphone data streams to build key figures for depression models?

Methods

Umbrella Review

Protocol and Registration

The umbrella review has been registered in the PROSPERO database (CRD42024581256) and was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [22].

Information Source and Search Strategy

During the preparation phase of the MENTINA trial in September 2024, the umbrella review was performed by querying the scientific databases PubMed, Web of Science Core Collection, and Scopus to get a comprehensive overview of the existing reviews in the field. The search strategy included querying the scientific databases with the search term "smartphone sensor data depression" to find reviews related to clinical depression or dealing with depression symptoms. To capture the most recent syntheses of evidence on smartphone data streams relevant for current mental health monitoring, filters were applied to include only review papers published within the past 5 years (2019-2024). This time frame aligns with findings that individuals typically replace their smartphones within a 5-year period [23], reinforcing the relevance of focusing on reviews that reflect contemporary device usage and capabilities.

Eligibility Criteria

Studies were considered only when the screening of the title and abstract confirmed that the study was a review written in English or German and focused on more than 1 smartphone sensor to build digital phenotypes of individuals' mental health. The authors JL and MP independently assessed the eligibility of the studies based on titles and abstracts, with UH serving as a supervising rater, resolving any conflicts that arose during the assessment process.

Selection of Studies

The full-text versions were obtained and the contents of potentially relevant papers were analyzed to select items matching the following criteria: (1) reviews already summarizing the use of raw smartphone sensors over multiple other works, and (2) reviews showing sensor purposes in terms of depression recognition. The preselection of studies based on the selection criteria was performed by author JL.

Quality Assessment

For quality assessment, reviews were evaluated using the AMSTAR (A Measurement Tool to Assess Systematic Reviews) checklist [24]. Reviews were chosen only if particular items could be answered with "yes". The items of relevance were the following:

- Item 2 ("Was there duplicate study selection and data extraction?")
- Item 3 ("Was a comprehensive literature search performed?")
- Item 6 ("Were characteristics of the included studies provided?")
- Item 11 ("Were potential conflicts of interest included?").

The full texts were independently quality-assessed by authors UH, MP, and JL. Final inclusion decisions were made through majority consensus.

Narrative Synthesis

To synthesize the relevant literature, each included review was screened for smartphone data streams previously used in mental health monitoring. These data streams were extracted and categorized into hardware (derived from installed components) and software (provided by the operating system) categories to distinguish their sources. For each identified data stream, the utilization for deriving mental health insights was assessed and extracted. The extracted data were organized into a table, listing smartphone data streams as rows and relevant studies as columns. Table cells indicated whether a data stream was mentioned in a study regarding mental health monitoring. If a data stream was mentioned but its specific use was unclear, the corresponding cell was marked as "listed." To enhance readability, similar terms were clustered, and a separate table mapped original terms to umbrella terms. The hardware data stream collection further served as the foundation for investigating cross-platform sensor availability.

The data extraction process was conducted by JL, with MP verifying the extracted data and decisions. UH supervised the process, resolving any conflicts.

Cross-Platform Sensor Availability

Smartphone Operating System Market Analysis

To assess the prevalence of commonly used mobile operating systems, market share data from the past 2 years was analyzed using statistics from the portal Statista [25-29]. To highlight the global significance of Android and iOS, worldwide market share data were extracted. Additionally, regional analyses were focused on Europe, with a particular emphasis on Denmark, Germany, and Spain, to provide insights into the operating system market landscape in these countries where the MENTINA trial will be conducted.

Additionally, to approximate real-world smartphone usage in Germany, the best-selling Android and iOS smartphones in the country for September 2024 were extracted from Counterpoint [30], and the sensor availability in these models was analyzed to better represent the likely scenarios in one of the MENTINA countries.

Phone Specifications Database

PhoneDB maintains the world's largest database for mobile device specifications, providing technical details for smartphones, tablets, smart watches, and several other types of handheld devices [31]. While PhoneDB is not a complete collection, it gives a comprehensive overview of smartphone releases over the past 20 years [32]. PhoneDB was used in a couple of other scientific works [32,33] and serves as a valuable

XSL•FO
RenderX

data source for deriving trends for technical specification trends. Besides information on the operating systems the devices are running, the database provides more than 290 parameters per device, including information on the presence of installed hardware sensors. PhoneDB offers licensed database dumps or free limited access to the database via preconfigured database queries using PhoneDB's web-based detailed parametric search tool. The search tool allows querying device specifications based on built-in sensors, which has been used for this work.

Data Queries

PhoneDB was queried multiple times to retrieve the number of devices matching the search criteria. The search criteria included filters on the device category, release date, operating system, and installed hardware including whether the devices are equipped with built-in hardware components. To identify operating system–dependent trends in hardware component availability, the database was queried for annual Android and iOS smartphone specifications with corresponding components installed from 2014 to 2024. To get a baseline, queries without any hardware component filter were executed to get the annual total number of Android and iOS releases for each year.

Additionally, PhoneDB was queried for each considered device to collect pricing information, including price, currency, market countries, and market region, to facilitate analysis of the correlation between price and hardware availability.

Data Analysis

The PhoneDB query results were saved in 1 file for iOS and 1 file for Android. Each file consisted of a data table with hardware components as rows and years as columns. The table cells represented the number of device releases in the particular year that had the corresponding hardware component installed. For the baseline, a row consisting of the total number of device releases over the considered years was added to each file. Based on the Android and iOS data tables, the hardware component availability as a percentage was calculated by dividing the number of device releases with the corresponding hardware component installed by the total number of device releases in the respective year and multiplying by 100.

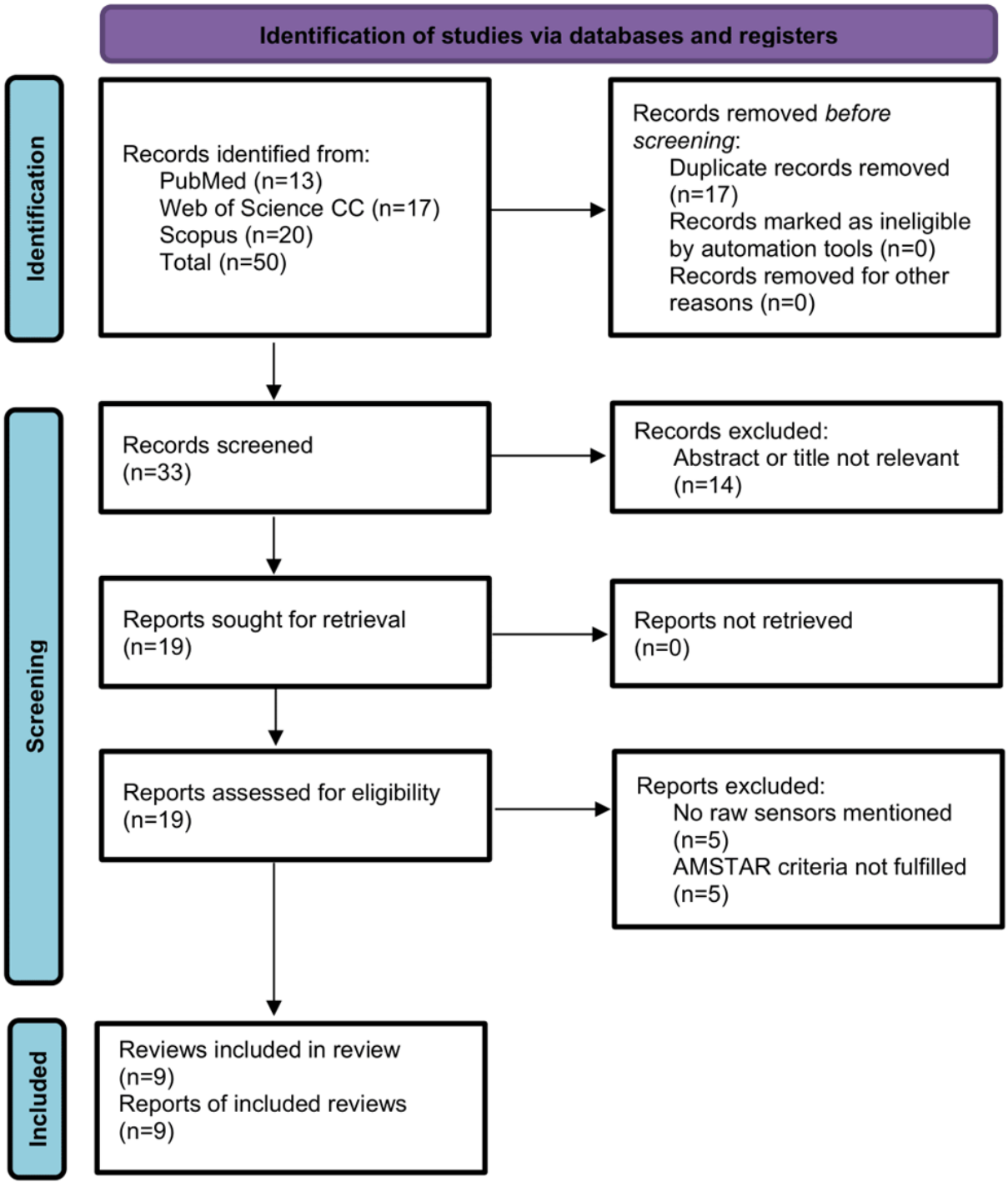
Results

Umbrella Review

A PRISMA flow chart (Figure 1) has been created to illustrate the information flow of the umbrella review. Aligned with the search strategy, 50 records were found in total across the 3 queried scientific databases (PubMed, n=13; Web of Science Core Collection, n=17; Scopus, n=20). After removing duplicates (n=17), 14 records were removed due to nonrelevant abstract and title. For the remaining 19 records, full-text reports were requested and screened to meet the eligibility criteria. Ten records were excluded for either not reporting the use of raw sensors (n=5) or for not meeting the AMSTAR criteria (n=5) [24]. Finally, 9 reviews were selected for data extraction and synthesis.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart. CC: Core Collection; AMSTAR: A Measurement Tool to Assess Systematic Reviews.



The studies included in the umbrella review are listed in Table 1. For transparency, the list of excluded records along with their exclusion reasons can be found in Multimedia Appendix 1. Table 2 depicts the results of the data extraction and synthesis by listing the found smartphone data streams in rows over the relevant studies as columns. The table cells indicate whether a smartphone data stream has been mentioned in the respective article regarding mental health monitoring.

XSL•FO
RenderX

Table 1. Included articles in the umbrella review.

Reference	Title
Beames et al [7], 2024	Use of smartphone sensor data in detecting and predicting depression and anxiety in young people (12-25 years): a scoping review
Trifan et al [8], 2019	Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations
Choi et al [9], 2024	Digital phenotyping for stress, anxiety, and mild depression: systematic literature review
Qirtas et al [10], 2022	Loneliness and social isolation detection using passive sensing techniques: scoping review
Yim et al [11], 2020	The utility of smartphone-based, ecological momentary assessment for depressive symptoms
Leaning et al [12], 2024	From smartphone data to clinically relevant predictions: a systematic review of digital phenotyping methods in depression
Virginia Anikwe et al [13], 2022	Mobile and wearable sensors for data-driven health monitoring system: state-of-the-art and future prospect
Lee et al [14], 2023	Using digital phenotyping to understand health-related outcomes: a scoping review
Gopalakrishnan et al [15], 2022	Mobile phone enabled mental health monitoring to enhance diagnosis for severity assessment of behaviours: a review

XSL•FO
RenderX

Table 2. Extracted smartphone data streams and their uses in mental health monitoring in each article included in the umbrella review (n=9).

Component	Reference								
	Beames et al [7]	Trifan et al [8]	Choi et al [9]	Qirtas et al [10]	Yim et al [11]	Leaning et al [12]	Virginia Anikwe et al [13]	Lee et al [14]	Gopalakrishnan et al [15]
Hardware data stream									
Accelerometer (n=9)	Physical activity	Physical activity, sleep, and well-being	Physical activity	Physical activity	Circadian rhythm	Physical activity	Physical activity and stress	Mobility, physical activity, and sleep	Physical activity and social activity
Barometer (n=1)	—[a]	Physical activity	—	—	—	—	—	—	—
Battery level (n=4)	—	Sleep, well-being, and circadian rhythm	Sleep	—	—	Listed[b]	—	Social activity	—
Bluetooth (n=8)	Social activity	Circadian rhythm, sleep, social activity, and well-being	Mobility and social activity	Social activity	Social activity	Environment	—	Mobility and social activity	Mobility and social activity
Camera (n=4)	—	Well-being	Listed	—	Listed	—	—	—	Social activity
Cellular network (n=3)	—	Listed	—	—	—	Mobility	—	—	Listed
GPS (n=9)	Mobility	Circadian rhythm, physical activity, sleep, and well-being	Mobility and physical activity	Mobility, physical activity, and social activity	Circadian rhythm	Circadian rhythm, environment, mobility, and physical activity	Circadian rhythm	Mobility	Mobility, physical activity, and social activity
Gyroscope (n=7)	Listed	Circadian rhythm, physical activity, social activity, and well-being	Physical activity and sleep	—	—	Mobility	Circadian rhythm	Physical activity	Physical activity
Humidity (n=1)	—	—	—	—	—	—	Well-being	—	—
Light (n=9)	Sleep	Listed	Sleep	Listed	Listed	Environment and sleep	Listed	Sleep	Sleep and social activity
Magnetometer (n=2)	—	Physical activity and well-being	—	—	—	—	Circadian rhythm	—	—
Proximity[c] (n=1)	—	—	—	—	—	—	Circadian rhythm	—	—

XSL•FO
RenderX

Component	Reference								
	Beames et al [7]	Trifan et al [8]	Choi et al [9]	Qirtas et al [10]	Yim et al [11]	Leaning et al [12]	Virginia Anikwe et al [13]	Lee et al [14]	Gopalakrishnan et al [15]
Sound (n=8)	Social activity	Circadian rhythm, physical activity, sleep, social activity, and well-being	Sleep and social activity	Sleep and social activity	Listed	—	Circadian rhythm	Social activity and sleep	Environment, sleep, and speech
Step count[d] (n=6)	Physical activity	Physical activity	Physical activity	—	—	Physical activity	Physical activity	Physical activity	—
Temperature (n=2)	—	—	—	—	—	—	Physical activity	—	Environment
Wi-Fi (n=8)	Mobility	Circadian rhythm, physical activity, sleep, social activity, and well-being	Mobility	Social activity	Listed	Mobility	—	Mobility	Mobility, physical activity, and social activity
Software data stream									
App usage (n=8)	Listed	Circadian rhythm, sleep, social activity, and well-being	Social activity	Social activity	Mood	Circadian rhythm, mood, and social activity	—	Social activity	Social activity and stress
Call and message logs (n=9)	Social activity	Circadian rhythm, sleep, social activity, and well-being	Social activity	Social activity	Mood	Social activity	Circadian rhythm	Social activity and stress	Social activity and stress
Screen (n=8)	Listed	Listed	Stress and well-being	Listed	Listed	Social activity	—	Social activity and sleep	Sleep

[a]The data stream was not mentioned.

[b]The data stream was mentioned but its specific use was unclear.

[c]Proximity sensors are usually realized with an emitting infrared LED and a detecting photodiode [34].

[d]Step counts are usually derived by sensor fusion of multiple other hardware sensors. Modern smartphones are equipped with dedicated motion coprocessors to do this task in an energy-efficient manner [35].

To improve the readability of Table 2, clusters were built for similar terms. A list of mapped original terms to umbrella terms is shown in Table 3.

XSL•FO
RenderX

Table 3. Mapping table of umbrella terms and assigned terms found across the literature.

Umbrella term	Original terms in the literature
Data stream	
App usage	app usage, application usage, browser usage, in-phone activity
Light	light, ambient light
Screen	screen status, lock/unlock status, screen time, screen events
Sound	sound, microphone, audio, voice, speech
Purpose	
Circadian rhythm	circadian rhythm, daily-life behavior, behavioral marker
Environment	environment, surrounding, workplace conditions
Mobility	mobility, location, homestay
Physical activity	physical activity, motion, human activity, movement
Sleep	sleep, bedtime, sleep disturbance
Social activity	social activity, loneliness, social avoidance, social rhythm, social function, social habits, social interaction, sociability, social anxiety
Stress	stress, distractedness

As the results in Table 2 show, the most mentioned hardware data streams included accelerometers, light sensors, and GPSs. While accelerometers are mostly used to derive a proxy measure for physical activity, light sensors are broadly used to gather insights about the sleep behavior of individuals. GPSs serve various purposes to build proxy measures for mental health [7-15]. The most prominent application of GPSs in terms of mental health monitoring is to derive information about individuals' mobility, which includes, for instance, the number of location changes or the time spent at home. Similar information could be also retrieved from the current position of a smartphone in cellular networks, which was also mentioned in 1 study [12]. In contrast to cellular network information, GPSs were also broadly used to track physical activity and helped to determine one's circadian rhythm, gathering, for instance, markers for daily life behavior. Only a few studies mentioned GPSs in context of measuring social activity [10,15], but according to the included studies, social activity is often approximated by identifying the number of closely located devices in one's environment using Bluetooth [7-11,14,15].

Similarly, some of the studies mention the use of microphones or cameras to get a measure for social activity by analyzing individuals' acoustic and visual environments [7-10,14,15]. Several studies also mention the use of microphones to estimate individuals' circadian rhythms, including sleep patterns [8-10,13-15]. Those patterns are also approximated by interpreting battery levels, as a couple studies referenced [8,9].

A couple of other hardware data streams were additionally mentioned for estimating circadian rhythms, including data derived from proximity sensors, gyroscopes, or Wi-Fi components [8,13]. Furthermore, Wi-Fi and gyroscopes were mentioned in the context of various other proxy measures for mental health monitoring, including mobility, social activity, and physical activity [7-10,12,14,15].

Beyond the hardware data streams already mentioned, step counts were used many times but exclusively to derive key figures for physical activity [7-9,12-14]. To complete the list of hardware data streams used for representing physical activity, temperature sensors, barometers, and magnetometers were also mentioned in the literature reviews for this purpose [8,13]. Finally, 1 relevant article stated the use of humidity sensors for approximating a person's well-being [13].

Furthermore, all included studies also listed software data streams for the monitoring of mental health (Table 2), including the data from call and message logs as an insightful data source for building proxy measures for mental health. While most studies stated that call and message logs were used to estimate individuals' social activity [7-10,12,14,15], 2 referred to using the logs for constructing metrics on circadian rhythm [8,13]. Additionally, estimates for mood and stress were listed as a derivate of call and message logs [11,14,15]. In principle, the same mental health characteristics were derived from using app and screen usage data across the literature, while a few studies stated that screen status was also used to approximate sleep behaviors [14,15].

Cross-Platform Sensor Availability

According to data from Statista [25], the worldwide smartphone operating system market is dominated by the Android and iOS operating systems, covering 72% and 28% of the market share, respectively, as of the second quarter of 2024. Similarly, the smartphone operating system market in Europe is dominated by Android and iOS [26]. In 2018, Android devices covered 67% of the European market, while the market share for iOS was 32%. In 2023, approximately 65% of European smartphones were still running Android, while the remaining market was almost completely covered by iOS devices. The mobile operating system market shares in Germany, which is part of the MENTINA trial, were similar to those of the broader European market, consisting of 60% Android devices and 39% iOS devices in 2023 [27]. As of March 2024, the mobile operating system market in Spain, another country included in the MENTINA trial, was dominated by Android with a share

XSL•FO
RenderX

of approximately 80%, while iOS covered approximately 20% [28]. In contrast, as of April 2024, 57% of mobile devices in Denmark, the third country in the MENTINA trial, were running iOS, while 42% of devices had Android installed [29]. Table 4 summarizes the operating system market shares for the considered regions.

Table 4. Smartphone operating system market shares according to Statista [25-29].

	Android (%)	iOS (%)
World	72	28
Europe	65	35
Germany	60	39
Spain	80	20
Denmark	57	42

To estimate real-world smartphone usage in Germany before the start of the MENTINA trial, the top 5 best-selling smartphones in the country for September 2024 were extracted from Counterpoint [30]. The Apple iPhone 15 Pro Max ranked first, accounting for 10% of sales, followed by the Apple iPhone 15 Pro (9%) and the Apple iPhone 16 Pro Max (6%). The list also included Android devices, with the Samsung Galaxy A55 5G ranking fourth (6%) and the Samsung Galaxy S24 ranking fifth (5%).

Due to the worldwide market dominance of Android and iOS smartphones, PhoneDB was only queried for hardware component availability on Android and iOS systems. Table 5 shows the fields and values of PhoneDB's search tool, combinations of which were used to estimate the trends of hardware component availability for Android and iOS smartphones over the past 10 years.

Table 5. Relevant PhoneDB fields and input values to meet the search criteria.

Search criteria	Field	Input values
General		
Year	Release Date	Dates spanning complete years from 2014 to 2024 (eg, from 2020-01-01 to 2020-12-31)
Smartphone	Device category	Smartphone
Operating system	Platform	Android or iOS/iPadOS
Data stream		
Accelerometer	Built-in accelerometer	Any
Barometer	Barometer checkbox	Checked
Battery	Battery	Any
Bluetooth	Bluetooth	Any
Camera	Camera image sensor	Any
Cellular network	Supported cellular bands	Any
GPS	GPS checkbox	Checked
Gyroscope	Built-in gyroscope	Any
Humidity	Humidity sensor checkbox	Checked
Light	L sensor checkbox	Checked
Magnetometer	Built-in compass	Any
Microphone	Microphone(s)	Any
Proximity	IR[a] face sensor or LiDAR[b] checkbox	Checked
Step counter	Step counter checkbox	Checked
Temperature	T sensor checkbox	Checked
Wi-Fi	Wireless LAN[c]	Any

[a]IR: infrared.

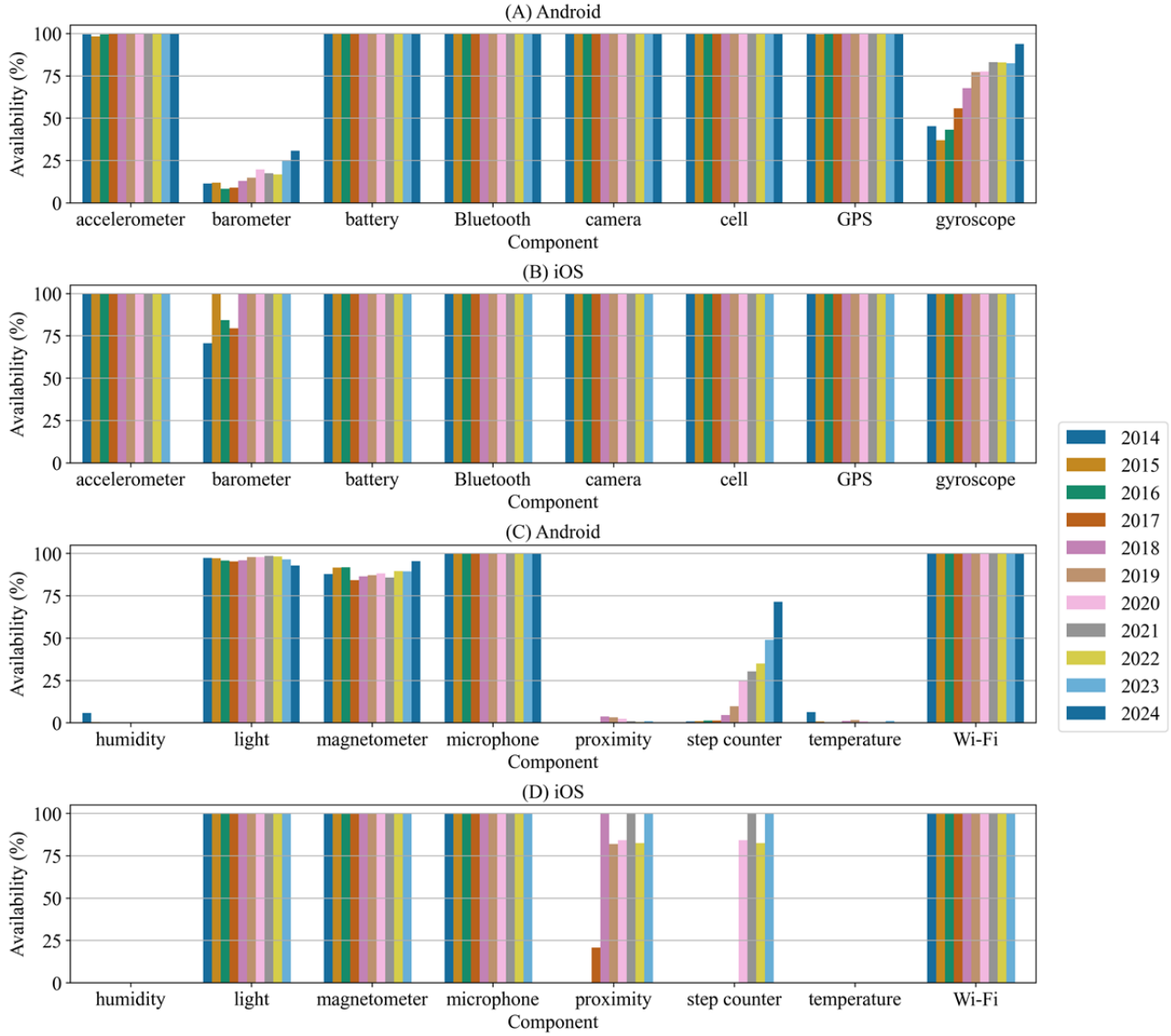
[b]LiDAR: light detection and ranging.

[c]LAN: local area network.

XSL•FO
RenderX

The raw data derived from the PhoneDB queries can be viewed in Multimedia Appendix 2. The appendix also contains the availability results for Android and iOS. Figure 2 depicts a condensed version of these results by showing the availability of 16 hardware components on Android and iOS smartphones from 2014 to 2024 as of September 2024.

Figure 2. Annual hardware component availability on Android and iOS smartphones from 2014 to 2024 as of September 2024. (A and B) Availability of accelerometer, barometer, battery Bluetooth, camera, cellular network, GPS, and gyroscope components on (A) Android and (B) iOS. (C and D) Availability of humidity, light, magnetometer, microphone, proximity, step counter, temperature, and Wi-Fi components on (C) Android and (D) iOS. The data for each hardware component are plotted in chronological ascending order to depict the availability trend throughout the years. The charts in B and D are missing a bar for 2024 because there were no existing iOS specifications for 2024 as of September 22, 2024.



As seen in Figure 2, hardware components such as accelerometers, batteries, Bluetooth, cameras, cellular network support, GPSs, microphones, and Wi-Fi were available on almost every Android and iOS smartphone released in the 10-year period. Since 2018, iOS smartphones have been equipped with barometers. In contrast, only 31% of Android devices came with barometers in 2024, but a positive trend was observed. Similarly, iOS smartphones have provided gyroscopes for the last 9 years, while not all Android devices were shipped with gyroscopes in 2024. With more than 80% of Android smartphones having gyroscopes installed from 2021 to 2023 and approximately 94% in 2024, a positive trend was visible. Besides a few recognizable Android devices in 2014, almost no Android or iOS smartphone provided a humidity sensor in the period studied. Almost the same pattern applied to temperature sensors, with several years seeing low availability on Android systems and no coverage on iOS systems. For light sensors, iOS devices provided full coverage from 2014 to 2023. For Android, 95% of devices provided light sensors from 2014 to 2023 and an all-time low of approximately 93% did so in 2024. Similarly, while magnetometers were found on every iOS smartphone, not every Android device came with magnetometers. The trend of magnetometer availability on Android phones experienced a slow increase from 2017 to 2023, reaching 95% availability in 2024. From 2018 to 2023, iOS smartphones came with a proximity sensor approximately 80% of the time, whereas almost no Android phones provided a proximity sensor during the studied period. Step counters are

XSL•FO
RenderX

usually derived by sensor fusion of multiple other hardware sensors, but modern smartphones are equipped with dedicated motion coprocessors to perform this task in a more energy-efficient manner [35]. Coprocessors for efficiently estimating step counts were present on iOS phones more than 80% of the time since 2017; Android systems have rapidly caught up, with an average availability of 71% in 2024.

For transparency, Multimedia Appendix 3 provides a comprehensive list of the smartphones considered in this study, including hardware component availability and price information sourced from PhoneDB. Among the top-selling iOS models, the Apple iPhone 15 Pro Max, iPhone 15 Pro, and iPhone 16 Pro Max include all hardware components listed in Table 5 except for humidity and temperature sensors. In contrast, the top-selling Android models, the Samsung Galaxy A55 5G and Galaxy S24, are equipped with fewer sensors than their iOS counterparts. The more expensive Samsung Galaxy S24 lacks sensors for measuring humidity, proximity, and temperature, while the most affordable among the 5 best-selling models, the Samsung Galaxy A55 5G, is missing additional components, including a barometer, humidity sensor, proximity sensor, step counter, and temperature sensor.

Discussion

Principal Results

To highlight the potential of sensors for building reliable mental health monitoring systems across smartphone platforms, we focused on answering three research questions in the course of this work: (1) Which data streams provided by smartphones have been used in the literature to build models for depression monitoring? (2) Which smartphone data streams are available across smartphone vendors and operating systems to build robust platform-independent software systems for depression monitoring? (3) Which proxies have been constructed from raw smartphone data streams to build key figures for depression models?

To address research question (1), an umbrella literature review aligned with the PRISMA workflow was conducted, and 9 articles were included that ensure high methodological quality according to the AMSTAR checklist. From each of the included articles, the mentioned smartphone data streams were extracted, resulting in a list of 16 hardware data streams and 3 software data streams.

To consider research question (3), the purposes of each data stream regarding mental health monitoring were extracted from the literature and assigned to similar purpose clusters. In summary, smartphone data were used in the literature to build key figures for circadian rhythm, the environment, mobility, mood, physical activity, sleep, social activity, speech, stress, and well-being to provide insights into the mental state of individuals.

To provide answers to research question (2), the identified smartphone hardware data streams were assessed regarding their availability on Android and iOS platforms.

In sum, hardware components including accelerometers, batteries, Bluetooth, cameras, cellular network receivers, GPSs, microphones, and Wi-Fi were available on every Android and iOS device released in the past decade. Due to their presence on every smartphone, data streams from these components are promising candidates for the construction of robust cross-platform software systems.

The same applies to data on gyroscopes, light sensors, and magnetometers, which were installed on every iOS and almost every Android smartphone. Despite the fact that step counts are already provided by the operating systems through sensor fusion of other hardware components, the availability of motion coprocessors for more accurate step counters on Android and iOS smartphones increased significantly throughout the studied period, making it another promising candidate for the establishment of robust software systems.

As of September 2024, barometers, humidity sensors, proximity sensors, and temperature sensors had not been consistently installed across Android and iOS mobile phones, and therefore have limited use for reliable cross-platform mental health monitoring systems.

Limitations

While this work strives to provide a complete picture of the smartphone data available for the development of reliable cross-platform depression indication systems, it does not consider the quality and uniformity of data streams. This work considers a data stream for a particular hardware component to be available if any type of hardware component has been installed. Due to the broad spectrum of installed hardware components in the considered time frame, the data stream signals will vary significantly, thus limiting the comparability and interpretability of raw values across mobile phones. Therefore, future work may focus on improving operating systems and middleware to provide comparable data streams derived from different hardware component types. Some studies have already attempted to address the integration of diverse sensor signals across mobile devices by providing cross-platform frameworks to facilitate the development of stable mobile apps [36,37].

Furthermore, this work lacks an in-depth analysis of how the extracted data streams were specifically used in the literature to compute key metrics for representing mental health. While the studies included in the umbrella review provide a high-level description of the use of smartphone data streams, more detailed research is needed for each data stream individually. An example of in-depth analysis is the investigation of human voice signals derived from smartphones for mood and depression monitoring, which plays a crucial role in the development of mental health monitoring systems [38,39]. Future research should further explore the preprocessing and integration of smartphone data streams to construct representative mental health metrics across platforms.

Although this study primarily focuses on the cross-platform availability of hardware sensors, the umbrella review also identified several software data streams relevant to mental health monitoring, including app usage, call and message logs, and screen status information. However, the included studies do not

XSL•FO
RenderX

provide a comprehensive overview of all available data streams offered by operating systems, as they do not account for data accessible through additional operating system functionalities or external application programming interfaces (APIs). For example, a study by Jacobson and Chung [40] illustrates how GPS location data can be enriched with weather conditions by retrieving temperature, humidity, and precipitation data from the National Weather Service API at a specific time and location. Given that humidity and temperature sensors are rarely integrated into Android and iOS smartphones, as indicated by our cross-platform sensor availability analysis, leveraging operating system internal or external APIs could help compensate for these hardware limitations by providing the necessary environmental data. Future research should investigate additional data streams that can be accessed via operating system APIs and external APIs, expanding the range of available information for mental health monitoring.

This umbrella review was designed to support the subsequent cross-platform availability analysis. Given the considerable redundancy observed across the included reviews, a broader search strategy would likely not have materially influenced the main findings, namely the smartphone data streams and their mental health purposes identified in recent years. Nevertheless, the comprehensiveness of the search strategy might be enhanced by the inclusion of additional keyword synonyms. In addition, the time frame for study inclusion was limited to the last 5 years, which may have led to the exclusion of earlier relevant reviews, particularly those covering foundational work or less frequently updated areas. Another limitation concerns the study selection process—while the selection criteria were clearly defined and straightforward to apply, full-text selection was performed by only 1 author (JL), which may have introduced a risk of selection bias. Furthermore, the umbrella review approach, which enabled a high-level synthesis of existing systematic reviews, comes with inherent limitations. As it depends on the scope and quality of the included reviews, any omissions or biases in those sources are carried forward. This may lead to underrepresentation of newer sensor types or recent primary studies not yet covered. Inconsistent reporting across reviews posed a challenge, but relevant information was systematically extracted and harmonized into a unified comparison table, enabling cross-review analysis of sensor modalities and mental health outcomes. Despite these constraints, the method provides a critically appraised foundation to inform future research and application design.

Comparison With Prior Work

Through the course of this work, plenty of reviews were found that already summarize the potential of smartphone data for mental health monitoring [7-15,41-44]. To our knowledge, there is no umbrella review providing a complete picture of the research in this field. Therefore, this work consolidates the literature findings with respect to the development of cross-platform mental health monitoring systems. The studies of the umbrella review [7-15] mostly highlight the potential of smartphone data for mental health monitoring, but little research has been performed on the availability of underlying data sources across mobile platforms. The availability of data streams is highly dependent on the smartphone operating system [42]. Therefore, this work closes that research gap by providing the availability of smartphone hardware data streams on the predominating smartphone platforms, Android and iOS, over the last 10 years. Vendors and researchers may make use of the availability results to build reliable and robust mental health monitoring software systems in the future.

Conclusions

Research in the field of digital phenotyping for mental health and depression modeling is an emerging and complex field that integrates insights from psychiatry, technology, and health informatics [45]. Currently, the literature lacks information on passive smartphone data for reliable key figures that represent mental health. This work has started to address this gap by analyzing the cross-platform availability of relevant smartphone data. The results are expected to support the informed decisions concerning the best data collection setup for apps and trials focusing on the value of long-term complex time series analyses of such data in people living with depression and other mental health issues. Nevertheless, further research is needed to streamline smartphone data across varying mobile hardware and software configurations to provide reliable digital mental health for everyone. Therefore, future work may investigate improving frameworks and operating systems to easily provide comparable data streams for mental health monitoring independent of the underlying smartphone.

Acknowledgments

This work was funded by the European Union (grant 101080651).

Authors' Contributions

JL conducted the cross-platform availability analysis, and the main writing of the manuscript. JL, MP and UH worked on the overall study design, conducted the umbrella review, and fine-tuned the manuscript. AD, DH, and UH supervised and reviewed the work from technical and medical perspectives.

Conflicts of Interest

None declared.

Multimedia Appendix 1

XSL•FO
RenderX

Studies excluded from the umbrella review and reasons for exclusion.
[DOCX File , 46 KB - ijmr_v14i1e69686_app1.docx]

Multimedia Appendix 2
The full version of the PhoneDB raw data and results of the cross-platform availability analysis.
[XLSX File (Microsoft Excel File), 15 KB - ijmr_v14i1e69686_app2.xlsx]

Multimedia Appendix 3
A comprehensive list of the considered mobile devices for each year including price information as provided by PhoneDB.
[XLSX File (Microsoft Excel File), 1676 KB - ijmr_v14i1e69686_app3.xlsx]

References

1. Kathan A, Harrer M, Küster L, Triantafyllopoulos A, He X, Milling M, et al. Personalised depression forecasting using mobile sensor data and ecological momentary assessment. Front Digit Health 2022;4:964582 [FREE Full text] [doi: 10.3389/fdgth.2022.964582] [Medline: 36465087]
2. Suhara Y, Xu Y, Pentland AS. DeepMood: forecasting depressed mood based on self-reported histories via recurrent neural networks. 2017 Presented at: 26th International Conference on World Wide Web; April 3-7, 20217; Geneva, Switzerland. [doi: 10.1145/3038912.3052676]
3. Lu J, Shang C, Yue C, Morillo R, Ware S, Kamath J, et al. Joint modeling of heterogeneous sensing data for depression assessment via multi-task learning. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-21. [doi: 10.1145/3191753]
4. Ware S, Yue C, Morillo R, Lu J, Shang C, Bi J, et al. Predicting depressive symptoms using smartphone data. Smart Health 2020 Mar;15:100093. [doi: 10.1016/j.smhl.2019.100093]
5. Masud MT, Mamun MA, Thapa K, Lee D, Griffiths MD, Yang S. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. J Biomed Inform 2020 Mar;103:103371 [FREE Full text] [doi: 10.1016/j.jbi.2019.103371] [Medline: 31935462]
6. He X, Triantafyllopoulos A, Kathan A, Milling M, Yan T, Rajamani S, et al. Depression diagnosis and forecast based on mobile phone sensor data. 2022 Presented at: 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society; July 11-15, 2022; Glasgow, Scotland. [doi: 10.1109/embc48229.2022.9871255]
7. Beames JR, Han J, Shvetcov A, Zheng WY, Slade A, Dabash O, et al. Use of smartphone sensor data in detecting and predicting depression and anxiety in young people (12-25 years): a scoping review. Heliyon 2024 Aug 15;10(15):e35472 [FREE Full text] [doi: 10.1016/j.heliyon.2024.e35472] [Medline: 39166029]
8. Trifan A, Oliveira M, Oliveira JL. Passive sensing of health outcomes through smartphones: systematic review of current solutions and possible limitations. JMIR Mhealth Uhealth 2019 Aug 23;7(8):e12649 [FREE Full text] [doi: 10.2196/12649] [Medline: 31444874]
9. Choi A, Ooi A, Lottridge D. Digital phenotyping for stress, anxiety, and mild depression: systematic literature review. JMIR Mhealth Uhealth 2024 May 23;12:e40689 [FREE Full text] [doi: 10.2196/40689] [Medline: 38780995]
10. Qirtas MM, Zafeiridi E, Pesch D, White EB. Loneliness and social isolation detection using passive sensing techniques: scoping review. JMIR Mhealth Uhealth 2022 Apr 12;10(4):e34638 [FREE Full text] [doi: 10.2196/34638] [Medline: 35412465]
11. Yim SJ, Lui LM, Lee Y, Rosenblat JD, Ragguett R, Park C, et al. The utility of smartphone-based, ecological momentary assessment for depressive symptoms. J Affect Disord 2020 Sep 01;274:602-609. [doi: 10.1016/j.jad.2020.05.116] [Medline: 32663993]
12. Leaning IE, Ikani N, Savage HS, Leow A, Beckmann C, Ruhé HG, et al. From smartphone data to clinically relevant predictions: a systematic review of digital phenotyping methods in depression. Neurosci Biobehav Rev 2024 Mar;158:105541 [FREE Full text] [doi: 10.1016/j.neubiorev.2024.105541] [Medline: 38215802]
13. Virginia Anikwe C, Friday Nweke H, Chukwu Ikegwu A, Adolphus Egwuonwu C, Uchenna Onu F, Rita Alo U, et al. Mobile and wearable sensors for data-driven health monitoring system: state-of-the-art and future prospect. Expert Syst Appl 2022 Sep;202:117362. [doi: 10.1016/j.eswa.2022.117362]
14. Lee K, Lee TC, Yefimova M, Kumar S, Puga F, Azuero A, et al. Using digital phenotyping to understand health-related outcomes: a scoping review. Int J Med Inform 2023 Jun;174:105061. [doi: 10.1016/j.ijmedinf.2023.105061] [Medline: 37030145]
15. Gopalakrishnan A, Venkataraman R, Gururajan R, Zhou X, Genrich R. Mobile phone enabled mental health monitoring to enhance diagnosis for severity assessment of behaviours: a review. PeerJ Comput Sci 2022;8:e1042 [FREE Full text] [doi: 10.7717/peerj-cs.1042] [Medline: 36092018]
16. Protecting mental health in times of change. Mentbest. URL: https://mentbest.com/ [accessed 2024-08-08]
17. 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 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

XSL•FO
RenderX

18. Arrieta J, Aguerrebere M, Raviola G, Flores H, Elliott P, Espinosa A, et al. Validity and utility of the Patient Health Questionnaire (PHQ)-2 and PHQ-9 for screening and diagnosis of depression in rural Chiapas, Mexico: a cross-sectional study. J Clin Psychol 2017 Sep;73(9):1076-1090 [FREE Full text] [doi: 10.1002/jclp.22390] [Medline: 28195649]
19. Monsenso. App Store. URL: https://apps.apple.com/us/app/monsenso/id1373639965 [accessed 2024-08-08]
20. Monsenso. Google Play. URL: https://play.google.com/store/apps/details?id=dk.monsenso.green&hl=de_AT [accessed 2024-08-08]
21. Monsenso. URL: https://www.monsenso.com/mhealth-mental-health/solution/ [accessed 2023-06-17]
22. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
23. Cordella M, Alfieri F, Clemm C, Berwald A. Durability of smartphones: a technical analysis of reliability and repairability aspects. J Clean Prod 2021 Mar 01;286:125388 [FREE Full text] [doi: 10.1016/j.jclepro.2020.125388] [Medline: 33658746]
24. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009 Oct;62(10):1013-1020. [doi: 10.1016/j.jclinepi.2008.10.009] [Medline: 19230606]
25. Market share of mobile operating systems worldwide from 2009 to 2024, by quarter. Statista. URL: https://www.statista.com/statistics/272698/global-market-share-held-by-mobile-operating-systems-since-2009/ [accessed 2024-11-07]
26. Market share of leading mobile operating systems in Europe from 2010 to 2023. Statista. URL: https://www.statista.com/statistics/639928/market-share-mobile-operating-systems-eu/ [accessed 2024-11-07]
27. Market share of mobile operating systems in Germany from 2010 to 2023. Statista. URL: https://www.statista.com/statistics/693829/market-share-mobile-operating-systems-germany/ [accessed 2024-11-07]
28. Market share of mobile operating systems in Spain from 2010 to 2024. Statista. URL: https://www.statista.com/statistics/640059/market-share-mobile-operating-systems-spain/ [accessed 2024-11-07]
29. Market share held by the leading mobile operating systems in Denmark as of April 2024. Statista. URL: https://www.statista.com/statistics/621147/most-popular-mobile-operating-systems-in-denmark/ [accessed 2024-11-07]
30. Top 5 smartphone models share for 8 countries. Counterpoint. URL: https://www.counterpointresearch.com/insight/top-5-smartphone-model-share-8-countries [accessed 2025-02-27]
31. PhoneDB. URL: https://phonedb.net/ [accessed 2024-05-16]
32. Jost T, Abou Baker N. Künstliche intelligenz ermöglicht automatisiertes smartphone-recycling. Prosperkolleg. URL: https://prosperkolleg.ruhr/wp-content/uploads/2022/09/rethink_22-05_smartphone-erkennung-mit-ki.pdf [accessed 2025-07-21]
33. Leguesse Y, Colombo C, Vella M, Hernandez-Castro J. PoPL: proof-of-presence and locality, or how to secure financial transactions on your smartphone. IEEE Access 2021;9:168600-168612. [doi: 10.1109/access.2021.3137360]
34. Rai A, Zhuo D, Bahreyni B. Passive proximity detection based on a miniaturized pyramidal optical sensor. 2019 Presented at: IEEE Sensors; October 27-30, 2019; Montreal, Canada. [doi: 10.1109/sensors43011.2019.8956834]
35. Lammel G, Dorsch R, Giesselmann T, Goldeck J, Hahn J, Hasan N, et al. Smart system architecture for sensors with integrated signal processing and AI. 2021 Presented at: Smart Systems Integration; April 27-29, 2021; Grenoble, France. [doi: 10.1109/ssi52265.2021.9466993]
36. Bardram J. The CARP mobile sensing framework -- a cross-platform, reactive, programming framework and runtime environment for digital phenotyping. ArXiv. Preprint posted online on June 21, 2020 . [doi: 10.48550/arXiv.2006.11904]
37. Langer P, Altmüller S, Fleisch E, Barata F. CLAID: Closing the Loop on AI & Data Collection — A cross-platform transparent computing middleware framework for smart edge-cloud and digital biomarker applications. Future Gener Comput Syst 2024 Oct 1;159:505-521. [doi: 10.1016/j.future.2024.05.026]
38. Flanagan O, Chan A, Roop P, Sundram F. Using acoustic speech patterns from smartphones to investigate mood disorders: scoping review. JMIR Mhealth Uhealth 2021 Sep 17;9(9):e24352 [FREE Full text] [doi: 10.2196/24352] [Medline: 34533465]
39. Shin J, Bae SM. Use of voice features from smartphones for monitoring depressive disorders: scoping review. Digit Health 2024;10:20552076241261920 [FREE Full text] [doi: 10.1177/20552076241261920] [Medline: 38882248]
40. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. Sensors (Basel) 2020 Jun 24;20(12):3572 [FREE Full text] [doi: 10.3390/s20123572] [Medline: 32599801]
41. Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. Sensors (Basel) 2019 May 09;19(9):2164 [FREE Full text] [doi: 10.3390/s19092164] [Medline: 31075985]
42. Kulkarni P, Kirkham R, McNaney R. Opportunities for smartphone sensing in e-Health research: a narrative review. Sensors (Basel) 2022 May 20;22(10):3893 [FREE Full text] [doi: 10.3390/s22103893] [Medline: 35632301]
43. Alamoudi D, Breeze E, Crawley E, Nabney I. The feasibility of using smartphone sensors to track insomnia, depression, and anxiety in adults and young adults: narrative review. JMIR Mhealth Uhealth 2023 Feb 17;11:e44123 [FREE Full text] [doi: 10.2196/44123] [Medline: 36800211]
44. Melcher J, Hays R, Torous J. Digital phenotyping for mental health of college students: a clinical review. Evid Based Ment Health 2020 Nov;23(4):161-166 [FREE Full text] [doi: 10.1136/ebmental-2020-300180] [Medline: 32998937]

XSL•FO
RenderX

45. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]

Abbreviations

AMSTAR: A Measurement Tool to Assess Systematic Reviews
API: application programming interface
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Wi-Fi: Wireless Fidelity

Edited by T de Azevedo Cardoso; submitted 19.12.24; peer-reviewed by P Domingues; comments to author 12.02.25; revised version received 27.03.25; accepted 15.07.25; published 07.08.25.

Please cite as:
Leimhofer J, Petrovic M, Dominik A, Heider D, Hegerl U
Cross-Platform Availability of Smartphone Sensors for Depression Indication Systems: Mixed-Methods Umbrella Review
Interact J Med Res 2025;14:e69686
URL: https://www.i-jmr.org/2025/1/e69686
doi:10.2196/69686
PMID:

©Johannes Leimhofer, Milica Petrovic, Andreas Dominik, Dominik Heider, Ulrich Hegerl. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 07.08.2025. 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

Glucagon-Like Peptide-1 Receptor Agonists Combined With Personalized Digital Health Care for the Treatment of Metabolic Syndrome in Adults With Obesity: Retrospective Observational Study

Hala Zakaria, MSc; Hadoun Jabri, MD; Sheikha Alshehhi, BSc; Milena Caccelli, MD; Joelle Debs, BSc; Yousef Said, MD; Joudy Kattan, MSc, MD; Noah Almarzooqi; Ali Hashemi, MBA; Ihsan Almarzooqi, PhD

GluCare.Health, Jumeirah 1, Alwasl road, Dubai, UAE, Dubai, United Arab Emirates

Corresponding Author:
Hala Zakaria, MSc
GluCare.Health, Jumeirah 1, Alwasl road, Dubai, UAE, Dubai, United Arab Emirates

Abstract

Background: Metabolic syndrome (MetS) represents a complex and multifaceted health condition characterized by a clustering of interconnected metabolic abnormalities, including central obesity, insulin resistance, dyslipidemia, and hypertension. Effective management of MetS is crucial for reducing the risk of cardiovascular diseases and type 2 diabetes.

Objective: This study aimed to assess the effectiveness of combining glucagon-like peptide-1 (GLP-1) and dual gastric inhibitory polypeptide (GIP)/GLP-1 agonists with a continuous, digitally delivered behavioral change model by an integrated care team, in treating MetS among individuals with obesity.

Methods: The 6-month Zone.Health (meta[bolic]) weight loss program involved 51 participants (mean age 45, SD 10 years; mean BMI 35, SD 5 kg/m²), categorized by gender, and treated with either tirzepatide or semaglutide. Participants received continuous support via a digital health platform, which facilitated real time monitoring and personalized feedback from an integrated care team. Engagement levels with the digital platform, measured by the frequency of inbound interactions, were tracked and analyzed in relation to health outcomes.

Results: Tirzepatide reduced waist circumference (WC) by −18.08 cm, compared with −13.04 cm with semaglutide (P<.001). Triglycerides decreased significantly with both drugs, with tirzepatide showing a reduction of −64.42 mg/dL and semaglutide −70.70 mg/dL (P<.001). Tirzepatide generally showed more pronounced improvements in fasting glucose, blood pressure (BP), low-density lipoprotein, and total cholesterol compared with semaglutide. Higher engagement with the digital health platform showed significant difference among the 3 groups; the group with the highest level of app-based interactions (≥25 interactions) had the greatest WC reduction (mean −19.04, SD 7.40 cm) compared with those with ≤15 interactions (mean −9.60, SD 5.10 cm; P=.002). Similarly, triglycerides showed the greatest reduction in the group with ≥25 interactions (mean −108.56, SD 77.06 mg/dL) compared with those with ≤15 interactions (mean −44.49, SD 50.85 mg/dL; P=.02). This group also exhibited the largest reduction in diastolic BP (mean −10.33, SD 7.40 mm Hg) compared with those with ≤15 interactions (mean −0.83, SD 7.83 mm Hg; P=.004), and the most substantial decrease in fasting glucose levels (mean −18.60, SD 10.82 mg/dL) compared with those with ≤15 interactions (mean −2.49, SD 27.54 mg/dL; P=.02). Participants in the highest quartile of digital engagement had a 60% greater likelihood of MetS reversal compared with those in the lowest quartile.

Conclusions: This study shows that combining GLP-1 and dual GIP/GLP-1 agonists with a digital behavioral change model significantly improves MetS markers in individuals with obesity. Tirzepatide proved more effective than semaglutide, leading to greater reductions in WC and triglyceride levels, along with better improvements in fasting glucose, BP, and lipid profiles. Higher app-based engagement was linked to better health outcomes, with participants in the highest engagement group having a 60% greater likelihood of treating MetS compared with those with the lowest engagement.

(Interact J Med Res 2025;14:e63079) doi:10.2196/63079

KEYWORDS

metabolic syndrome; obesity; GLP-1 medications; hybrid model of care; digital health; effectiveness; digital engagement; hybrid care; adult; cardiovascular disease; type 2 diabetes; insulin resistance; efficacy; behavioral change; obese; zone health; weight loss; monitoring; tirzepatide; semaglutide; treatment; medication; telehealth; health informatics; glucagon-like peptide-1

XSL•FO
RenderX

Introduction

Metabolic syndrome (MetS) represents a complex and multifaceted health condition characterized by a clustering of interconnected metabolic abnormalities, including central obesity, insulin resistance, dyslipidemia, and hypertension [1]. Individuals diagnosed with MetS face an elevated risk of developing cardiovascular disease [2], type 2 diabetes [3], and other related complications [4]. Understanding the intricate mechanisms underlying MetS is crucial for optimizing patient management and outcomes. In the United Arab Emirates, the prevalence of MetS was 33.6% (269/801) in the Emirati population, 34.5% (214/620) in the Arab non-Emirati population, and 40.7% (695/1709) in the Asian non-Arab population [5].

Obesity serves as a primary contributing factor to the development and progression of MetS. Excess adipose tissue, particularly visceral fat, contributes to chronic low-grade inflammation and the release of adipokines, cytokines, and free fatty acids, all of which play vital roles in insulin resistance and metabolic dysregulation [6]. Central obesity, characterized by an accumulation of adipose tissue around the abdomen, is particularly indicative of MetS and poses a heightened risk for cardiovascular complications [7]. The management of MetS revolves around lifestyle modifications and pharmacological interventions aimed at addressing its components and reducing overall cardiovascular risk [8]. Lifestyle interventions, including dietary changes, regular physical activity, and weight management, form the cornerstone of treatment [9]. In patients with MetS, dietary strategies emphasizing a balanced intake of macronutrients and lifestyle modification such as smoking cessation, regular exercise, and proper eating habits may improve profiles of each component of MetS and reduce the risk of developing diabetes and cardiovascular disease [9].

Pharmacotherapy plays a complementary role in the management of MetS, with medications targeting specific components of the syndrome. This may include statins to address dyslipidemia, antihypertensive agents to manage elevated blood pressure (BP), and insulin-sensitizing drugs to improve glucose metabolism [8]. Notably, emerging therapeutic agents such as glucagon-like peptide-1 (GLP-1) receptor agonists have shown promise in addressing multiple aspects of MetS, including glycemic control, weight reduction, and cardiovascular risk reduction [10]. By mimicking the action of endogenous GLP-1, these medications stimulate glucose-dependent insulin secretion, suppress glucagon release, and delay gastric emptying, resulting in improved glycemic control and reduced appetite [10]. Furthermore, GLP-1 receptor agonists have been associated with favorable effects on body weight, BP, and lipid profiles, making them attractive therapeutic options for individuals with MetS [10].

MetS poses a significant challenge to public health, necessitating comprehensive and more engaging approaches for prevention and management. Integrating dietary modifications, lifestyle interventions, and pharmacological treatments targeting specific components of the syndrome are essential for mitigating its adverse outcomes. The Zone.Health's meta[bolic] program, which combines pharmacotherapy with continuous engagement and monitoring to enable sustainable lifestyle modifications, demonstrated significant improvements in weight, body composition, and metabolic markers [11]. Enhancing traditional treatment paradigms by integrating continuous digital health monitoring with tailored pharmacotherapy. This program leverages advanced analytics and real time data to dynamically adjust treatment plans, fostering deeper patient engagement and more precise management of MetS components. Compared with traditional models that often rely on intermittent follow-ups and generalized treatment approaches, Zone.Health's continuous care model ensures long-term lifestyle modifications and pharmacological adherence, which are critical for long-term management of MetS. This integration of technology and personalized care is designed to significantly improve clinical outcomes by providing consistent, supportive, and adaptive interventions tailored to individual patient needs. Therefore, the aim of this study is to measure the effectiveness of GLP-1 and GLP-1/gastric inhibitory polypeptide (GIP) medications when combined with a continuous, digitally delivered behavioral change model involving a multidisciplinary team in the treatment of MetS components and weight among individuals with obesity in the UAE population.

Methods

Zone.Health Program

Zone.Health, introduced by meta[bolic] [12] in 2023, is a value-based 6-month weight loss program designed to be used in combination with pharmacotherapy. This innovative initiative offers hyperpersonalized insights, and monitoring in order to have patients adopt realistic, bite-sized behavioral change decisions over the duration of the program. The program's "value-based" approach underscores its commitment to efficacy, promising partial refunds to participants who fulfill specified compliance metrics but do not achieve a minimum 10% weight loss within the program's timeframe. This ensures accountability while prioritizing the participant's progress and well-being. Hyperpersonalization lies at the core of Zone.Health, where each participant's unique physiological, behavioral, and psychological profile is considered. When focusing on physiological factors, parameters like age, fat mass, muscle mass, and physical activity levels are considered. Whereas the behavioral and psychological aspect includes mental health, goal setting, and psychological beliefs. Furthermore, using continuous glucose monitoring (CGM), alongside activity and sleep tracking, and other digital biomarkers, the health care team dynamically tailors dietary recommendations, exercise plans, and medication regimens to suit each participant needs. The ongoing engagement between participants and the health care team is necessary. Biweekly sessions by the health coach comprehensive provides support, addressing progress, challenges, and evolving needs. As participants advance, engagement frequency adjusts to monthly sessions, focusing on maintaining achievements and adapting strategies for long-term success. Zone.Health's multifaceted approach extends beyond face-to-face interactions, incorporating an intuitive app called "Zone.Health" supported by the health care team including coaches, sports scientists, dietitians, physicians, and

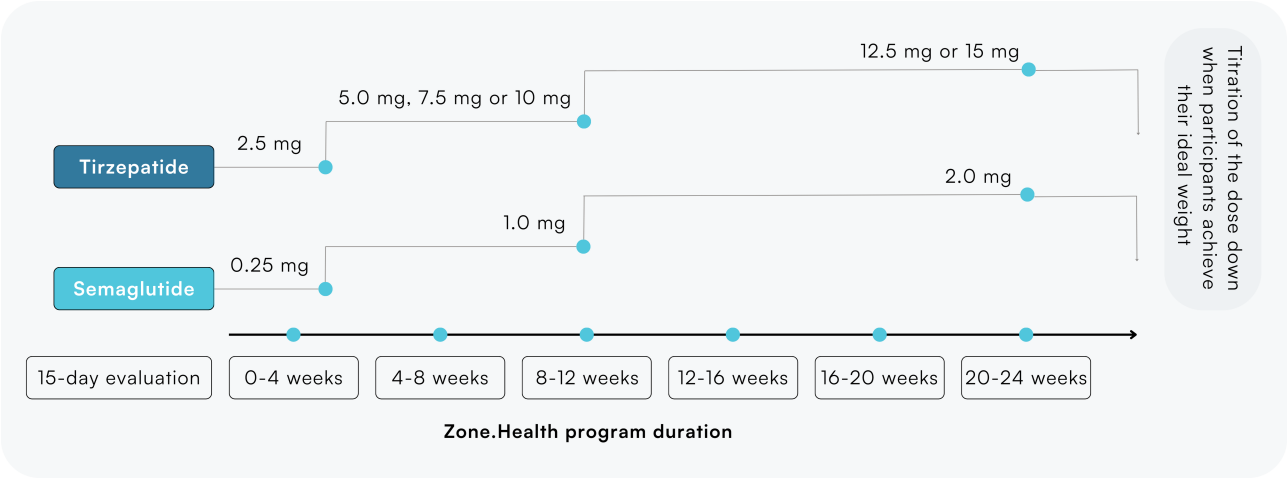
XSL•FO
RenderX

diabetes educators. The app is integrated with a nutrition artificial intelligence food logging tool used by participants for food logging as well as integration with wearable data from Apple healthkit or Google fit. Participants are educated by dietitians to ensure correct food logging and minimize errors, by diabetes educators to interpret CGM data, and by health coaches to ensure integrating all other digital wearables (including Apple watch, Google Fitbit, or ŌURA ring [Oura Health Ltd]) into the app and educate the participants data collected from these wearables. Data integration is seamless, with unified and time-synchronized information accessible through a single clinician portal, facilitating effective patient monitoring and engagement. Medication adjustments are made monthly, guided by assessments of fat versus muscle loss ratio, weight fluctuations, and reported side effects. Regular measure of body composition analysis using (Seca mBCA 514, Seca GmbH & Co. KG) ensures a comprehensive understanding of each participant's progress, with evaluations conducted both at the program's onset and quarterly thereafter for comparison. The effectiveness of the Zone.Health's hybrid approach is supported by previous research, particularly in diabetes management, demonstrating its potential to revolutionize weight loss strategies and improve overall health outcomes [11,13,14].

Study Design

A retrospective, real world evidence observational study was conducted among participants with MetS that signed up to the Zone.Health weight loss program. A comprehensive examination of medical records was undertaken to explore the occurrence and factors contributing to MetS among patients who were prescribed GLP-1 medications. This study encompassed 51 adult patients from diverse ethnic backgrounds who had completed a 6-month program duration with all MetS parameters collected. Medication administration followed a thorough assessment, including a 15-day evaluation period and initial dietary analysis, to confirm participant eligibility. Treatment selection, whether semaglutide or tirezepatide, was determined by physicians based on clinical suitability. Following the standard practice, tirzepatide was started with an initial dose of 2.5 mg, gradually increasing to either 5.0 mg, 7.5 mg, or 10 mg by the 3-month mark and reaching 12.5 mg or 15 mg by 6 months. Semaglutide started at 0.25 mg, escalating to 1 mg at 3 months and reaching to 2.0 mg at 6 months (Figure 1). All participants adhered to the criteria, ensuring the continued validity of contractual agreements. The study strictly followed the ethical guidelines of the Declaration of Helsinki, ensuring each participant gave signed consent. The UAE health care authority monitored the clinical protocols, ensuring they met international ethical standards for research involving human subjects. All study procedures, including data collection and analysis, respected participant rights and privacy.

Figure 1. Tirzepatide and semaglutide dose titration throughout the Zone.Health program.



Study Outcomes

Primary Objective

Reduction in MetS markers (waist circumference [WC], triglycerides, and fasting glucose levels) among individuals with obesity over a 6-month period when treated with GLP-1 (semaglutide) and dual GIP/GLP-1 agonists (tirzepatide).

Secondary Objectives

First, changes in BP (systolic and diastolic) and lipid profiles (total cholesterol, low-density lipoprotein [LDL], and high-density lipoprotein [HDL]) in participants receiving tirzepatide versus those receiving semaglutide.

Second, association between levels of digital engagement with the health platform and improvements in MetS markers.

Criteria for Diagnosing MetS

International Diabetes Federation (IDF) definition of MetS was used accordingly to classify participants with MetS [15]. In this study, females and males were included in the study if they had the MetS if the following criteria were met.

First, waist circumference of ≥94 cm for males and ≥80 cm for females, based on ethnic-specific values, such as Europids, Sub-Saharan Africans, Ethnic South and Central Americans, and Middle Eastern individuals.

Second, any two of the following four factors: (1) raised triglycerides ≥150 mg/dL (1.7 mmol/L) or specific treatment

XSL•FO
RenderX

for this lipid abnormality; (2) reduced HDL cholesterol (males <40 mg/dL [1.03 mmol/L], females <50 mg/dL [1.29 mmol/L]), or specific treatment for this lipid abnormality; raised blood pressure (systolic BP ≥130 mm Hg or diastolic BP ≥85 mm Hg), or treatment for previously diagnosed hypertension; (4) raised fasting plasma glucose ≥100 mg/dL (5.6 mmol/L), or previously diagnosed type 2 diabetes.

Data Collection and Participants

Data were collected from physicians' patient records at baseline and at 6 months using the electronic medical record (Diamond, Hicom). Patients' gender, age, ethnicity, medication, WC (cm), weight kg, BP (mm Hg), and BMI (kg/m²) variables were collected. Laboratory variables were collected including total cholesterol (mg/dL), LDL cholesterol (mg/dL), HDL cholesterol (mg/dL), triglycerides (mg/dL), and fasting glucose (mg/dL). Engagement interactions were collected from the app portal. Inbound interactions refer to interactions or messages received from the participants related to dietary advice or modifications, medication dosage and side effects, coaching-related queries, and inquiries directed at any of the health care team. The outbound interactions refer to the interactions or messages initiated by the health care team to the participants, which consists of (1) biweekly reminders of food logging, weight recording, data integration, and monthly workout schedules (dietitians, coaches, and sport scientists); (2) biweekly physicians follow-up on the side effect of the medications; and (3) CGM data feedback provided by the diabetes educator.

Participant Inclusion Criteria

Eligible participants were adults (aged 18 years and older) diagnosed with MetS and obesity, who met the IDF criteria for MetS, which includes central obesity (WC specific to population and gender) plus at least 2 of the following: raised triglycerides, reduced HDL cholesterol, raised BP, or raised fasting plasma glucose. Participants had been prescribed GLP-1 or GIP/GLP-1 therapies as part of their treatment regimen. Individuals with incomplete records or those who did not engage with the digital platform were excluded from the analysis.

Ethical Considerations

Ethical approval was obtained from the Dubai Health Authority (DHA; DSREC-02/2025_09), and patient data were anonymized to maintain confidentiality. To protect participant privacy, all data collected were anonymized, and measures were taken to safeguard sensitive information, including data encryption and restricted access. It is important to note that no compensation was provided to participants for their involvement in this study.

Statistical Analysis

Data analysis performed using SPSS software version 29.0 (IBM Corp). Continuous data (such as age, weight, and laboratory parameters) were presented as means and SDs, and categorical variables (such as gender, medication type, Total n° of IDF criteria met, and ethnicity) were expressed as counts and percentage. The paired *t* test was used to compare variables at baseline and at 6 months and ANOVA test was used to see the baseline characteristics between female and male. One-way ANOVA test was conducted to test differences in baseline characteristics across medication types. ANOVA was also used to assess differences between app-based interaction groups and their impact on improvements in MetS markers. $P<.05$ was considered significant. Furthermore, post hoc Bonferroni test was conducted to measure the difference between app-based engagement groups.

Results

Basic Demographics and Baseline Characteristics Stratified by Sex (N=51)

The study population is almost evenly divided between individuals of Middle Eastern (23/51, 45%) and European (28/51, 55%) ethnicities (Table 1 and Table 2). Majority of participants are treated with tirzepatide (31/51, 61%) as opposed to semaglutide (20/51, 39%). In terms of the number of MetS parameters met, all patients met the IDF criteria with most participants (34/51, 67%) meeting 3 criteria, 27% (14/51) meeting 4, and a small group of 6% (3/51) meeting all 5 criteria. Differences between the tirzepatide and semaglutide groups are evident in several baseline characteristics. There is a significant difference in age, with the tirzepatide group being older ($P=.02$). Notable disparities also exist in weight and WC, with the tirzepatide group being significantly heavier ($P=.01$) and having a larger WC ($P=.002$). The tirzepatide group also has a higher BMI ($P=.02$). BP measurements between the groups are similar, without significant differences in either systolic or diastolic BP. Cholesterol profiles are comparable, with no significant differences in total cholesterol, LDL cholesterol, or HDL cholesterol levels. In addition, there are no significant differences in triglycerides or fasting glucose levels between the 2 groups.

XSL•FO
RenderX

Table . Basic demographics and baseline characteristics in all participants classified as having the metabolic syndrome according to the IDF[a] criteria.

Characteristic		Values (N=51), n (%)
Ethnicity		
	Middle Eastern	23 (45)
	European	28 (55)
Sex		
	Female	28 (55)
	Male	23 (45)
Total number of IDF criteria met		
	3	34 (67)
	4	14 (27)
	5	3 (6)

[a]IDF: International Diabetes Federation.

Table . Clinical and metabolic characteristics of participants on tirzepatide and semaglutide.

Characteristics	Tirzepatide (n=31), mean (SD)	Semaglutide (n=20), mean (SD)	*P* value
Age (years)	47.03 (10.49)	40.05 (9.61)	.02[a]
Weight (kg)	105.90 (26.60)	88.61 (14.23)	.01[a]
BMI (kg/m^2)	35.43 (6.90)	31.49 (3.04)	.02[a]
Systolic BP[b] (mm Hg)	127.52 (14.66)	123.55 (13.00)	.33
Diastolic BP (mm Hg)	79.03 (8.90)	76.80 (7.82)	.36
Total cholesterol (mg/dL)	198.11 (46.69)	204.52 (40.03)	.62
LDL[c] cholesterol (mg/dL)	138.39 (43.91)	149.25 (45.42)	.40
Waist circumference (cm)	114.26 (17.64)	100.33 (9.72)	.002[a]
HDL[d] cholesterol (mg/dL)	47.51 (10.61)	50.93 (11.40)	.28
Triglycerides (mg/dL)	156.41 (70.21)	171.41 (83.93)	.50
Fasting glucose (mg/dL)	106.02 (10.16)	101.72 (22.43)	.38

[a]*P* values <.05 were considered significant from 1-way ANOVA test.

[b]BP: blood pressure.

[c]LDL: low-density lipoprotein.

[d]HDL: high-density lipoprotein.

Improvements of Diagnostic Criteria of MetS After 6 Months of Zone.Health Program

Table 3 illustrates significant improvements in the criteria used to diagnose MetS after a 6-month program, comparing the effects of tirzepatide and semaglutide (N=51). Tirzepatide showed greater reductions in WC (–18.08 cm) compared with semaglutide (−13.04 cm), both statistically significant (*P*<.001). Triglyceride levels decreased significantly with tirzepatide (−64.42 mg/dL) and even more with semaglutide (−70.70 mg/dL; *P*<.001). HDL cholesterol increases were modest and not statistically significant for both drugs. Significant reductions were observed in fasting glucose, systolic and diastolic BP, and LDL cholesterol, with tirzepatide generally showing more pronounced improvements than semaglutide. Overall cholesterol levels also decreased significantly in both treatment groups.

XSL•FO
RenderX

Table . Improvements of diagnostic criteria of metabolic syndrome after 6 months of the program in both female and male (N=51).

Variables	Tirzepatide (n=31)			Semaglutide (n=20)		
	Mean (SD)	*t* test (df)	*P* value	Mean (SD)	*t* test (df)	*P* value
Weight (kg)	−14.07 (5.83)	−13.44 (30)	<.001[a]	−13.38 (4.69)	−12.75 (19)	<.001[a]
BMI (kg/m^2)	−4.96 (2.54)	−10.69 (29)	<.001[a]	−4.53 (1.89)	−10.69 (19)	<.001[a]
Waist circumference (cm)	−18.08 (8.30)	−12.13 (30)	<.001[a]	−13.04 (7.27)	−8.02 (19)	<.001[a]
Triglycerides (mg/dL)	−64.42 (53.10)	−6.44 (29)	<.001[a]	−70.70 (69.10)	−4.58 (19)	<.001[a]
HDL[b] cholesterol (mg/dL)	2.50 (8.47)	1.68 (30)	.10	1.97 (8.67)	1.01 (19)	.32
Fasting glucose (mg/dL)	−15.32 (11.46)	−7.07 (27)	<.001[a]	−10.25 (23.64)	−1.89 (18)	.07
Systolic BP[c] (mm Hg)	−14.74 (13.92)	−5.90 (30)	<.001[a]	−14.75 (10.76)	−6.12 (19)	<.001[a]
Diastolic BP (mm Hg)	−7.23 (9.20)	−4.37 (30)	<.001[a]	−6.00 (8.32)	−3.22 (19)	.004
LDL[d] cholesterol (mg/dL)	−24.46 (32.58)	−4.04 (28)	<.001[a]	−15.15 (23.64)	−1.86 (18)	.08
Cholesterol (mg/dL)	−28.87 (45.02)	−3.51 (29)	.001[a]	−20.09 (39.98)	−2.13 (17)	.05[a]

[a]*P* values <.05 were considered significant from paired sample *t* test.

[b]HDL: high density lipoprotein.

[c]BP: blood pressure.

[d]LDL: low-density lipoprotein.

Treatment of Remaining MetS Markers in Patients With Increased WC (n=26)

In participants who did not reverse MetS due to persistently increased WC, improvements were observed with the remaining MetS markers. For instance, 85% (22/26) of participants successfully treated their elevated triglyceride levels to within normal ranges. HDL saw a significant improvement with 73% (19/26) of participants elevating their levels to the desired range. In addition, BP was stabilized in 89% (23/26) of these participants, and blood sugar levels were brought back to healthy levels in 96% (25/26) of those who did not treat MetS due to persistent increased WC.

Change in MetS Risk Factors by App-Based Interaction Groups

Table 4 and Table 5 presents the difference between patient engagement and improvements in MetS markers over a 6-month period among all participants. Engagement is quantified as both inbound and outbound interactions; inbound interactions (received from participants) average at 28.86 (SD 21.32), outbound (sent to participants from the health care team) at 58.57 (SD 25.26), and the total number of interactions at 84.00 (SD 45.76). Significant improvements were observed in several markers. WC, triglycerides, diastolic BP, and fasting glucose showed significant changes, with *P* values of .003, .01, .003, and .03, respectively. Changes in HDL and systolic BP were profound but not statistically significant. The post hoc analysis showed significant differences among interaction groups, indicating that higher app-based interactions are associated with better outcomes (Table 6). A significantly increased reduction in WC was found in individuals with ≥25 interactions (mean −19.04, SD 7.40) compared with individuals with ≤15 interactions (mean −9.60, SD 5.10 cm, *P*=.002). In addition, individuals with ≥25 interactions showed a significant reduction in triglycerides (mean −108.56, SD 77.06 mg/dL) compared with those with ≤15 interactions (mean −44.49, SD 50.85 mg/dL, *P*=.02) and a significant reduction when compared with those with 16 - 24 interactions (mean −55.77, SD 44.52 mg/dL, *P*=.03). Diastolic BP also significantly decreased in individuals with ≥25 interactions (mean −10.33, SD 7.40 mm Hg) compared with those with ≤15 interactions (mean −0.83, SD 7.83 mm Hg, *P*=.004). Fasting glucose levels were significantly lower in individuals with ≥25 interactions (mean −18.60, SD 10.82 mg/dL) compared with those with ≤15 interactions (mean −2.49, SD 27.54 mg/dL, *P*=.02).

XSL•FO
RenderX

Table . Distribution of engagement interactions by category.

Engagement interactions and category	Mean (SD)
Inbound	28.86 (21.32)
Outbound	58.57 (25.26)
Total	84.00 (45.76)

Table . Change in metabolic syndrome risk factors by app-based interaction groups (N=51).

Improvements in MetS[a] markers	Inbound app-based interactions			
	≤15 (n=12)	16-24 (n=12)	≥25 (n=27)	*P* value
WC[b] (cm)	−9.60 (5.10)	−16.01 (9.20)	−19.04 (7.40)	.003[c]
Triglycerides (mg/dL)	−44.49 (50.85)	−55.77 (44.52)	−108.56 (77.06)	.01[c]
HDL[d] (mg/dL)	+1.69 (10.03)	+2.93 (7.33)	+3.15 (5.63)	.85
Systolic BP[e] (mm Hg)	−12.58 (10.39)	−14.33 (12.79)	−15.89 (13.78)	.76
Diastolic BP (mm Hg)	−0.83 (7.83)	−4.58 (9.29)	−10.33 (7.40)	.003[c]
FG[f] (mg/dL)	−2.49 (27.54)	−13.41 (8.88)	−18.60 (10.82)	.03[c]

[a]MetS: metabolic syndrome.

[b]WC: waist circumference.

[c]*P* values <.05 were considered significant from 1-way ANOVA test.

[d]HDL: high-density lipoprotein.

[e]BP: blood pressure.

[f]FG: fasting glucose.

XSL•FO
RenderX

Table . Post hoc analysis of metabolic syndrome markers by interaction group (N=51).

Metabolic syndrome markers and interaction group (I)	Interaction group (J)	Mean difference (I–J)	*P* value
WC[a] (cm)			
≤15	16 - 24	6.41	.12
≤15	≥25	9.44[b]	.002[b]
16 - 24	≤15	−6.42	.12
16 - 24	≥25	3.03	.73
≥25	≤15	−9.44[b]	.002[b]
≥25	16 - 24	−3.03	.73
Triglycerides (mg/dL)			
≤15	16 - 24	11.28	≥.99
≤15	≥25	64.07[b]	.02[b]
16 - 24	≤15	−11.28	≥.99
16 - 24	≥25	52.79[b]	.03[b]
≥25	≤15	64.07[b]	.02[b]
≥25	16 - 24	−52.79[b]	.03[b]
Diastolic BP[c] (mm Hg)			
≤15	16 - 24	3.75	.76
≤15	≥25	9.50[b]	.004[b]
16 - 24	≤15	−3.75	.76
16 - 24	≥25	5.75	.13
≥25	≤15	−9.50[b]	.004[b]
≥25	16 - 24	−5.75	.13
FG[d] (mg/dL)			
≤15	16 - 24	10.92	.35
≤15	≥25	16.10[b]	.02[b]
16 - 24	≤15	−10.92	.35
16 - 24	≥25	5.19	≥.99
≥25	≤15	−5.19	≥.99
≥25	16 - 24	−16.10[b]	.02[b]

[a]WC: waist circumference.

[b]Mean difference is considered significant at .05 from Bonferroni post hoc test.

[c]BP: blood pressure.

[d]FG: fasting glucose.

Discussion

Principal Findings

This study examines the effect of GLP-1 and dual GIP/GLP-1 agonists when combined and integrated with a tech-enhanced, continuous feedback cycle on lifestyle modifications in treating MetS. MetS is intricately linked to central obesity and excessive weight, with studies illustrating that increased adiposity, particularly in the abdominal region, is a major risk factor for developing MetS [16]. Central obesity is associated with increased insulin resistance, which plays a pivotal role in the pathogenesis of MetS [16].

Studies have shown that even a modest amount of weight loss is associated with improvement in MetS parameters and related cardiovascular disease risk components. Losing weight is associated with a decrease in insulin resistance and insulin level, decreasing the risk of type 2 diabetes or improving previously diagnosed diabetes [17]. Losing weight also causes a positive impact on dyslipidemia by an increase in the HDL level and a drop in the LDL level [18,19]. GLP-1 and GIP agonist use has

XSL•FO
RenderX

been shown to reduce MetS severity, abdominal obesity, and inflammation [20].

Over a 6-month period, the comparative analysis of tirzepatide and semaglutide in the Zone.Health program showed significant improvements in MetS markers. Tirzepatide led to greater reductions in WC and LDL cholesterol, with slightly better results in fasting glucose and BP than semaglutide. While both treatments modestly increased HDL cholesterol without significance, semaglutide had a slightly better effect on lowering triglycerides. This suggests the treatment choice may vary based on individual patient needs and goals.

Despite significant improvements in WC observed in the study, it is noted that the rate of change might be perceived as low relative to other MetS markers. This could be attributed to several factors. First, central obesity, characterized by high WC, is often more stubborn and resistant to reduction through conventional weight loss measures alone [21]. This resistance could be partly due to the physiological characteristics of adipose tissue in the abdominal area, which is more metabolically active and sensitive to hormonal influences [21]. This may be attributed to the physiological characteristics of abdominal adipose tissue, which is highly metabolically active and hormonally sensitive, making it particularly challenging to reduce through pharmacotherapy and lifestyle changes alone. In addition, the IDF criteria consider an increased WC as a mandatory component for the diagnosis of MetS, which could preserve the perceived low reversal rate in study participants despite other significant improvements. The other MetS markers such as triglycerides, fasting glucose, and BP showed more substantial improvements even with persistent increased WC. These markers often respond more dynamically to metabolic improvements induced by GLP-1 agonists and lifestyle interventions. GLP-1 agonists improve insulin sensitivity and secretion, which directly impacts glucose control and lipid metabolism [18,19]. This improvement in insulin dynamics could explain the significant decrease in triglycerides and fasting glucose [22]. Similarly, the improvements in BP may be due to the weight loss itself as well as the improvement in arterial stiffness and endothelial function resulting from better metabolic control [23]. Furthermore, the baseline WC of our participants was significantly higher than the general thresholds for MetS, suggesting that longer-term or more intensive treatment strategies might be necessary to achieve meaningful reductions in WC. Challenges in reducing WC in this study align with findings from the Look AHEAD trial, where intensive lifestyle interventions resulted in significant weight loss but less pronounced reductions in WC [24]. This discrepancy highlights the complexity of central obesity as a metabolic entity and underscores the need for targeted strategies that specifically address this component of MetS.

Integrating Digital Solutions in Weight Loss Management and MetS Markers

The integration of hybrid care models in chronic disease management, such as diabetes mellitus, has shown promising results in reducing key metrics like HbA1c levels [11,13,14]. Our study also explores the perspective of people with MetS receiving treatment with medication such as semaglutide or tirzepatide along with feedback relating to their health in a real world environment. The long-term treatment of obesity, diabetes, and other chronic diseases requires that health care professionals shift away from episodic care and keep patients continuously engaged. The integration of digital tools as part of a standard care pathway allows health care practices to monitor and communicate with patients by observing personal data received from patients, whether it be activity, sleep habits, or information on dietary choices. Together, they are collectively able to take action regarding medication titration or a more consolidated plan during treatment [25]. The association between patient engagement and improvements in MetS markers, as presented in Table 3 of our study, further underscores the potential of digital health solutions in chronic disease management, particularly when combined with pharmacotherapy. Our findings indicate that higher engagement in inbound interactions is linked to better outcomes in key MetS parameters such as WC, triglycerides, and diastolic BP, and are supported by previous studies around engagement and outcomes in diabetes [13]. The significant reduction in these MetS markers with increased digital interaction suggests that patient engagement may be as critical as medication adherence in the management of MetS. This aligns with the broader trend in health care that emphasizes patient-centered models where ongoing patient engagement and personalized interaction play pivotal roles. The integration of digital tools facilitates this by providing continuous monitoring and real time feedback, which are essential for sustaining behavior changes over time. Furthermore, the Peterson Report emphasizes that purely digital solutions often fall short in achieving long-term patient engagement and effective disease management, citing a lack of personalization and direct human interaction as key shortcomings [26]. This finding supports the adoption of a hybrid care model, which combines the scalability of digital tools with the tailored support and expertise of the treating health care professionals.

To our knowledge, this study is the first to examine the effectiveness of GLP-1 medications alongside lifestyle interventions under a hybrid model of care in improving MetS markers. However, this study has a few limitations. First, the sample size, although adequate for detecting significant changes, is not large enough to ensure the generalizability of the findings across diverse populations or to detect more subtle effects. Second, the absence of a control group makes it difficult to definitively attribute the improvements in MetS markers to the interventions, as comparisons with a nonintervention baseline are lacking. In addition, Zone.Health is a paid program, and monetary investments into the program may have influenced participant outcomes, as those who have paid for the program might be more likely to participate actively and follow instructions. What can be concluded is that GLP outcomes on certain MetS parameters can be enhanced when delivered in a hybrid care model, as demonstrated in previous studies, thereby leading to better outcomes, an important factor considering the cost of GLP treatments.

Conclusions

In conclusion, this study highlights the effective role of combining GLP-1 and GIP/GLP-1 agonists with a continuous,

XSL•FO
RenderX

digitally delivered behavioral change model by an integrated care team in managing MetS, particularly in reducing WC, improving lipid profiles, and enhancing glycemic control across different genders. Despite notable improvements in most MetS markers, the persistence of increased WC suggests a need for strategies specifically targeted at abdominal obesity. Future research should focus on optimizing interventions that specifically address this resistant aspect of MetS, potentially incorporating new digital biomarker monitoring to enhance engagement even further due to the relationship between engagement and outcomes, as shown in this study.

Data Availability

The datasets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HZ, HJ, JK, and IA contributed to the conceptualization. HZ, SA, NA, and IA contributed to the methodology. HZ, SA, HJ, JD, and IA contributed to the investigation. Project administration was handled by HZ. HZ and IA contributed to the visualization. HZ and SA contributed to data curation. HZ, HJ, SA, and IA contributed to the writing – original draft. MC, JD, YS, JK, NA, and AH contributed to the writing – review and editing.

Conflicts of Interest

The following authors are full-time employees or interns at GluCare: HZ, HJ, SA, MC, JD, YS, JK, and NA. The following authors have affiliations with organizations with direct or indirect financial interest in the subject matter discussed in the manuscript: AH and IA.

References

1. Han TS, Lean MEJ. Metabolic syndrome. Medicine (Abingdon) 2015 Feb;43(2):80-87. [doi: 10.1016/j.mpmed.2014.11.006]
2. Guembe MJ, Fernandez-Lazaro CI, Sayon-Orea C, Toledo E, Moreno-Iribas C, RIVANA Study Investigators. Risk for cardiovascular disease associated with metabolic syndrome and its components: a 13-year prospective study in the RIVANA cohort. Cardiovasc Diabetol 2020 Nov 22;19(1):195. [doi: 10.1186/s12933-020-01166-6] [Medline: 33222691]
3. Shin JA, Lee JH, Lim SY, et al. Metabolic syndrome as a predictor of type 2 diabetes, and its clinical interpretations and usefulness. J Diabetes Investig 2013 Jul 8;4(4):334-343. [doi: 10.1111/jdi.12075] [Medline: 24843675]
4. Alberti K, Eckel RH, Grundy SM, et al. Harmonizing the metabolic syndrome: a joint interim statement of the International Diabetes Federation Task Force on Epidemiology and Prevention; National Heart, Lung, and Blood Institute; American Heart Association; World Heart Federation; International Atherosclerosis Society; and International Association for the Study of Obesity. Circulation 2009 Oct 20;120(16):1640-1645. [doi: 10.1161/CIRCULATIONAHA.109.192644] [Medline: 19805654]
5. Mahmoud I, Sulaiman N. Prevalence of metabolic syndrome and associated risk factors in the United Arab Emirates: a cross-sectional population-based study. Front Public Health 2021;9:811006. [doi: 10.3389/fpubh.2021.811006] [Medline: 35141192]
6. Kawai T, Autieri MV, Scalia R. Adipose tissue inflammation and metabolic dysfunction in obesity. Am J Physiol Cell Physiol 2021 Mar 1;320(3):C375-C391. [doi: 10.1152/ajpcell.00379.2020] [Medline: 33356944]
7. Chait A, den Hartigh LJ. Adipose tissue distribution, inflammation and its metabolic consequences, including diabetes and cardiovascular disease. Front Cardiovasc Med 2020;7:22. [doi: 10.3389/fcvm.2020.00022] [Medline: 32158768]
8. Rask Larsen J, Dima L, Correll CU, Manu P. The pharmacological management of metabolic syndrome. Expert Rev Clin Pharmacol 2018 Apr;11(4):397-410. [doi: 10.1080/17512433.2018.1429910] [Medline: 29345505]
9. Kim HL, Chung J, Kim KJ, et al. Lifestyle modification in the management of metabolic syndrome: statement from Korean Society of CardioMetabolic Syndrome (KSCMS). Korean Circ J 2022 Feb;52(2):93-109. [doi: 10.4070/kcj.2021.0328] [Medline: 35128848]
10. Seon MJ, Hwang SY, Son Y, Song J, Kim OY. Circulating GLP-1 levels as a potential indicator of metabolic syndrome risk in adult women. Nutrients 2021 Mar 6;13(3):865. [doi: 10.3390/nu13030865] [Medline: 33800785]
11. Zakaria H, Alshehhi S, Caccelli M, et al. Effectiveness of a hybrid approach in integrating GLP-1 agonists and lifestyle guidance for obesity and pre-diabetes management: RWE retrospective study. Metabol Open 2024 Jun;22:100283. [doi: 10.1016/j.metop.2024.100283] [Medline: 38699398]
12. Meta[bolic]. URL: https://metabolic.health/ [accessed 2025-03-24]
13. Zakaria H, Said Y, Aleabova S, et al. Measuring the effectiveness of hybrid diabetes care over 90 days through continuous data monitoring in type 2 diabetic patients. Front Endocrinol (Lausanne) 2024;15:1355792. [doi: 10.3389/fendo.2024.1355792] [Medline: 38774233]
14. Almarzooqi I, Zakaria H, Aleabova S, et al. Effectiveness of a hybrid technology enabled care model as measured by ICHOM standard set on established and managed type 2 diabetes already using medications: a RWE retrospective study. Metabol Open 2023 Dec;20:100262. [doi: 10.1016/j.metop.2023.100262] [Medline: 38115869]

XSL•FO
RenderX

15. Zimmet P, M M Alberti KG, Serrano Ríos M. A new international diabetes federation worldwide definition of the metabolic syndrome: the rationale and the results. Rev Esp Cardiol 2005 Dec;58(12):1371-1376. [Medline: 16371194]
16. Grundy SM. Metabolic syndrome pandemic. Arterioscler Thromb Vasc Biol 2008 Apr;28(4):629-636. [doi: 10.1161/ATVBAHA.107.151092] [Medline: 18174459]
17. Yamakage H, Jo T, Tanaka M, et al. Five percent weight loss is a significant 1-year predictor and an optimal 5-year cut-off for reducing the number of obesity-related cardiovascular disease risk components: the Japan obesity and metabolic syndrome study. Front Endocrinol (Lausanne) 2024;15:1343153. [doi: 10.3389/fendo.2024.1343153] [Medline: 38601201]
18. Williams PT, Stefanick ML, Vranizan KM, Wood PD. The effects of weight loss by exercise or by dieting on plasma high-density lipoprotein (HDL) levels in men with low, intermediate, and normal-to-high HDL at baseline. Metab Clin Exp 1994 Jul;43(7):917-924. [doi: 10.1016/0026-0495(94)90277-1] [Medline: 8028519]
19. Khan AA, Mundra PA, Straznicky NE, et al. Weight loss and exercise alter the high-density lipoprotein lipidome and improve high-density lipoprotein functionality in metabolic syndrome. Arterioscler Thromb Vasc Biol 2018 Feb;38(2):438-447. [doi: 10.1161/ATVBAHA.117.310212] [Medline: 29284607]
20. Sandsdal RM, Juhl CR, Jensen SBK, et al. Combination of exercise and GLP-1 receptor agonist treatment reduces severity of metabolic syndrome, abdominal obesity, and inflammation: a randomized controlled trial. Cardiovasc Diabetol 2023 Feb 25;22(1):41. [doi: 10.1186/s12933-023-01765-z] [Medline: 36841762]
21. Després JP, Lemieux I. Abdominal obesity and metabolic syndrome. Nature New Biol 2006 Dec 14;444(7121):881-887. [doi: 10.1038/nature05488] [Medline: 17167477]
22. Nauck MA, Meier JJ, Cavender MA, Abd El Aziz M, Drucker DJ. Cardiovascular actions and clinical outcomes with glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors. Circulation 2017 Aug 29;136(9):849-870. [doi: 10.1161/CIRCULATIONAHA.117.028136] [Medline: 28847797]
23. Landsberg L, Aronne LJ, Beilin LJ, et al. Obesity-related hypertension: pathogenesis, cardiovascular risk, and treatment--a position paper of the The Obesity Society and The American Society of Hypertension. Obesity (Silver Spring) 2013 Jan;21(1):8-24. [doi: 10.1002/oby.20181] [Medline: 23401272]
24. Dutton GR, Lewis CE. The Look AHEAD trial: implications for lifestyle intervention in type 2 diabetes mellitus. Prog Cardiovasc Dis 2015;58(1):69-75. [doi: 10.1016/j.pcad.2015.04.002] [Medline: 25936906]
25. Rhee SY, Kim C, Shin DW, Steinhubl SR. Present and future of digital health in diabetes and metabolic disease. Diabetes Metab J 2020 Dec;44(6):819-827. [doi: 10.4093/dmj.2020.0088] [Medline: 33389956]
26. Digital Diabetes Management Solutions: Peterson Health Technology Institute (PHTI); 2024.

Abbreviations

BP: blood pressure
CGM: continuous glucose monitoring
GIP: gastric inhibitory polypeptide
GLP-1: glucagon-like peptide-1
HDL: high-density lipoprotein
IDF: International Diabetes Federation
LDL: low-density lipoprotein
MetS: metabolic syndrome
WC: waist circumference

Edited by TDA Cardoso; submitted 10.06.24; peer-reviewed by AJ Graybeal, L Qian; revised version received 20.02.25; accepted 24.02.25; published 27.03.25.

Please cite as:
Zakaria H, Jabri H, Alshehhi S, Caccelli M, Debs J, Said Y, Kattan J, Almarzooqi N, Hashemi A, Almarzooqi I
Glucagon-Like Peptide-1 Receptor Agonists Combined With Personalized Digital Health Care for the Treatment of Metabolic Syndrome in Adults With Obesity: Retrospective Observational Study
Interact J Med Res 2025;14:e63079
URL: https://www.i-jmr.org/2025/1/e63079
doi:10.2196/63079

© Hala Zakaria, Hadoun Jabri, Sheikha Alshehhi, Milena Cacceli, Joelle Debs, Yousef Said, Joudy Kattan, Noah Almarzooqi, Ali Hashemi, Ihsan Almarzooqi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.3.2025. 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,

XSL•FO
RenderX

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

Interactive Effects of Weight Recording Frequency and the Volume of Chat Communication With Health Care Professionals on Weight Loss in mHealth Interventions for Noncommunicable Diseases: Retrospective Observational Study

Yuta Hagiwara[1], MSci; Takuji Adachi[2], PhD; Masashi Kanai[1,3], PhD; Kotoe Shimizu[1], BSc; Shinpei Ishida[1], MSci; Takahiro Miki[1], PhD

[1]PREVENT Inc, IKKO Shinsakae Building, 1 Aoi, Higashi-ku, Aichi, Japan

[2]Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Aichi, Japan

[3]Institute of Transdisciplinary Sciences for Innovation, Kanazawa University, Kanazawa, Japan

Corresponding Author:
Takahiro Miki, PhD
PREVENT Inc, IKKO Shinsakae Building, 1 Aoi, Higashi-ku, Aichi, Japan

Abstract

Background: Mobile health (mHealth) apps are increasingly used for health promotion, particularly for managing noncommunicable diseases (NCDs) through behavior modification. Understanding the factors associated with successful weight loss in such interventions can improve program effectiveness.

Objective: This study examined factors influencing weight change and the relationship between weight recording frequency and chat volume with health care professionals on weight loss in individuals with obesity and NCDs.

Methods: The participants had obesity (BMI ≥25 kg/m²) and were diagnosed with NCDs (eg, hypertension, diabetes, dyslipidemia). The program included 12 telephone consultations with health care professionals. Only participants who completed the full 6-month program, including all 12 telephone consultations, and provided an end-of-study weight were included in the analysis. The primary outcome was the rate of weight change, defined as the percentage change in weight from the initial period (first 14 days) to the final period (2 weeks before the last consultation), relative to the initial weight. The key independent variables were proportion of days with weight recording and chat communication volume (total messages exchanged). An interaction term between these variables was included to assess moderation effects in the regression analysis. The volume of communication was measured as the total number of messages exchanged, with each message, regardless of who sent it, being counted as 1 interaction. Health care staffs were instructed to send a single scheduled chat message per week following each biweekly phone consultation. These scheduled messages primarily included personalized feedback, reminders, and motivational support. In addition, providers responded to participant-initiated messages at any time during the program. Furthermore, 1 professional responded to each participant. Hierarchical multiple regression and simple slope analyses were conducted to identify relationships and interactions among these variables.

Results: The final analysis of this study included 2423 participants. Significant negative associations were found between the rate of weight change and baseline BMI (β=−.10; *P*<.001), proportion of days with weight recording (β=−.017; *P*<.001), and communication volume (β=−.193; *P*<.001). The interaction between proportion of days with weight recording and chat frequency also showed a significantly negative effect on weight change (β=−.01; *P*<.001). Simple slope analysis showed that when the proportion of days with weight recording was +1 SD above the mean, frequent chats were associated with greater weight reduction (slope=−0.60; *P*<.001), whereas no significant effect was observed at −1 SD (slope=−0.01; *P*=.94)

Conclusions: The findings suggest that both the proportion of days with weight recording and communication volume independently and interactively influence weight change in individuals with obesity and NCDs.

(Interact J Med Res 2025;14:e65863) doi:10.2196/65863

KEYWORDS

weight change; behavior modification; health care communication; weight recording; chat communication; text communication; health care professionals; weight loss; mHealth; mobile health; app; digital health; smartphone; mobile health intervention; noncommunicable disease; NCD; weight loss outcome; obesity; overweight; retrospective study; observational study; cerebrovascular disease; cardiovascular disease; lifestyle modification; mobile phone

XSL•FO
RenderX

Introduction

Obesity is a major public health challenge and a well-established risk factor for noncommunicable diseases (NCDs), such as hypertension, diabetes, and dyslipidemia. Effective weight management is crucial not only for obesity prevention but also for reducing the risk of NCD-related complications. Obesity is recognized as one of the contributing factors to the onset of NCDs, such as ischemic heart disease, cerebrovascular disease, hypertension, and diabetes mellitus. It was found that 43% of individuals aged 18 years and older were overweight and 16% were living with obesity in 2022 over the world [1,2]. Several guidelines for primary and secondary prevention of NCDs recommend weight loss for managing risk factors derived from insulin resistance [3-6]. Interventions causing any weight loss significantly reduce systolic and diastolic blood pressure, low-density lipoprotein cholesterol, triglycerides, fasting plasma glucose, and HbA_{1c} in a period of over 6 - 12 months, and lasting benefits persist for at least 2 years [7]. Furthermore, other studies have shown that patients with diabetes who achieved ≥10% weight loss in the first year after diagnosis had a significantly higher likelihood of remission [8]. Thus, effective strategies for weight reduction are important for managing and preventing NCDs.

The use of mobile health (mHealth) apps is an emerging approach, which has been proven to promote weight loss through lifestyle modifications [9]. One of the key functions of mobile apps is daily lifestyle records and health status monitoring, which empower individuals to actively manage their health and enhance efficiency for health care providers [10]. These apps offer tools for self-monitoring, allowing users to record their weight, dietary intake, and physical activity. Study has confirmed the effectiveness of self-monitoring in promoting weight loss [11], with frequent weight recording being particularly associated with greater weight loss [12,13].

Recently, interactive approaches that include communication with health care professionals are considered highly effective in mHealth intervention [14,15]. These involve phone consultations, chat messaging, and personalized feedback, supporting swifter information exchange between the user and health care professionals. Studies have also shown that personalized advice and social support from health care professionals enhance adherence to weight loss programs and improve outcomes [14,15]. The success of these mHealth interventions is due to user engagement, specifically, the frequency of users recording their data in the app. High engagement rates are crucial for effective self-monitoring and allow health care professionals to provide relevant feedback.

Various opinions exist on what constitutes the most effective form of interactive support within these programs. Although some argue for the quality of interactions, such as the content and personalization of messages, others have focused on the importance of the quantity of interactions. Our review of the literature suggests that the volume of communication between users and health care professionals could be a critical factor in the effectiveness of weight loss interventions [16-20]. However, only limited evidence can confirm the interactive effect of adherence to self-monitoring and the quantity of app-based communication volume on achieving weight reduction.

Therefore, this study aims to investigate the relation of weight recording frequency and app-based communication volume with weight reduction and determine their interactive effects on the participants of the app-based lifestyle modification.

Methods

Study Design and Setting

This retrospective observational study used data from PREVENT Inc., a company providing medical data analysis and mobile app-based lifestyle modification support programs for individuals with NCDs. PREVENT Inc. predicts the risk of the onset of cerebrovascular and cardiovascular diseases using health insurance claims and health checkup data. Furthermore, employees or their dependents who are at high risk of onset could be provided with an individualized support program by health care professionals.

Ethical Considerations

This study was conducted in accordance with the Declaration of Helsinki, and was approved by the Research Ethics Committee of Nagoya University (approval 2022 - 0453). Participants agreed to a privacy policy at the start of the program, which stated that the data gathered in the app may be used for future study. All participants provided informed consent before participating in the study. No financial compensation was provided to participants, who were free to withdraw at any time. All data presented were anonymized, and there are no images in the paper or the Multimedia Appendix that could identify individual participants.

Study Population and Procedure

PREVENT, Inc provides a mobile app-based lifestyle modification support program called Mystar to individuals at a high risk of cerebrovascular and cardiovascular diseases with NCDs, commissioned by the health insurance association. The selection of the target population was performed objectively using health insurance claims and health checkup data provided by the health insurance association. The principal approach to participation in these targeted population programs was for the health insurance association or PREVENT, Inc to send an invitation to participate in the program, or for the health insurance association to contact the individual. This study analyzed data from individuals who participated in the program from December 2018 to November 2023.

Participants eligible for this study were required to meet specific inclusion criteria. These criteria included a diagnosis of hypertension, diabetes, or dyslipidemia, with participants either receiving medication or having a history of coronary artery disease or stroke. Furthermore, participants needed to be classified as having obesity, defined by a BMI of 25 kg/m^2 or higher. Only individuals who agreed to participate in the 6-month program and provided informed consent for the use of their data in study were included. Exclusion criteria were applied to ensure the integrity of the study. Participants who failed to complete the full 6-month program (dropouts) or whose BMI

XSL•FO
RenderX

was below 25 kg/m² at the time of enrollment were excluded. In addition, individuals with medical conditions that could interfere with participation, such as severe cardiac diseases (eg, arrhythmia or cardiomyopathy), end-stage kidney disease, or mental disorders, were not eligible. The study also excluded participants who were taking medications that might affect weight management interventions, such as cardiotonic agents, immunosuppressive drugs, or antipsychotics. Finally, participants deemed unsuitable for inclusion by the health insurance association were excluded based on predefined criteria.

Disease Management Program

The disease management program was implemented with the approval of the participants' attending physicians. The program's objective was to improve the management of certain conditions, such as hypertension, diabetes, and dyslipidemia, through lifestyle improvements, to prevent the onset of conditions, such as stroke and myocardial infarction. Participants were provided with a mobile app account, which allowed them to record their lifestyle habits and exchange chat messages with health care professionals assigned to them. The program lasted 6 months, during which the participants had 12 telephone consultations with health care professionals. Nurses, registered dietitians, physical therapists, and other health care professionals provided individual advice through phone calls and chats based on the participants' lifestyle and disease management conditions. During telephone consultations, the participants set behavioral goals for the next 2 weeks and worked on improving lifestyle habits, such as exercise and diet by recording self-monitoring of health behaviors. Using the chat feature, not only could the participants ask health care professionals questions anytime but health care professionals could also provide individual advice. Further details of the program are described in a previous study [21].

Measurement of the Study Outcome: Weight Change

This study's primary outcome was that the rate of weight loss because the rate of weight change has been shown to be important for health improvement in individuals with obesity in Japan [22]. The rate of weight loss was calculated as the percentage change in weight over the program duration. Specifically, it was determined by dividing the difference between the initial average weight (recorded during the first 14 days of the program) and the final average weight (recorded during the 2 weeks before the last consultation) by the initial average weight, and then multiplying by 100 to express it as a percentage. This approach ensures that weight change is represented as a proportion of the initial weight, rather than just an absolute difference.

Proportion of Days With Weight Recording

Weight recording was measured as the proportion of days during the program on which participants recorded their weight at least once. If a participant recorded their weight more than once on a single day, only the first entry was used to calculate daily adherence. Units are reported as the percentage of days with weight recording relative to the total number of days in the program.

Communication Volume

Communication volume was assessed by counting the total number of chat messages exchanged between participants and health care staff throughout the program. Each chat message, whether initiated by the participants or health care staff, was considered as 1 interaction. In addition, 1 professional responded to each participant. The participants received advice from their assigned health care staff a week after each of the biweekly phone consultations, which was implemented over 6 months, totaling 12 sessions. Chat interactions included responses to participant inquiries, personalized feedback, and reminders for weight tracking when participants failed to log their weight. Health care professionals involved in these interactions included nurses, registered dietitians, and physical therapists, and messages were tailored based on participants' conditions and progress. Apart from these scheduled messages, the participants and health care professionals could freely engage in chat exchanges at any time during the program. In these chat communications, the health care staff provided feedback, lifestyle modification guidance, and information that encouraged continued healthy behaviors based on the participants' self-monitoring of health behaviors data. The participants could report the details of their daily activities and ask the health care staff about any questions or concerns. Each chat message, whether initiated by the participants or health care staff, was considered as 1 interaction.

Other Variables

We collected comprehensive demographic data (eg, age, gender, height, weight, and BMI), medical information, self-monitoring of health behaviors data, and app usage statistics from the participants. In addition, medical history information was obtained from health insurance claims data before the program began. Demographic and medication data were collected at program enrollment, while self-monitoring and app usage data were recorded throughout the program.

Independent Variables

The key independent variables in this study were the proportion of days with weight recording and chat communication volume. In addition, age, gender, and BMI were included as independent variables to control for demographic and physiological factors. An interaction term between the proportion of days with weight recording and chat communication volume was introduced in Model 2 to examine their combined effect on weight change.

Statistical Analysis

To handle missing data, a listwise deletion method was used, ensuring that only complete cases were analyzed. Descriptive statistics were used to summarize the basic demographic and clinical characteristics of the study population, including age, gender, BMI, proportion of days with weight recording, and volume of communication with health care professionals. To assess potential selection bias, we compared baseline characteristics between participants who completed the program and those who were excluded due to missing data. Independent *t* tests were used for continuous variables, and χ^2 tests were applied for categorical variables. A hierarchical multiple regression analysis was performed to investigate the relation

XSL•FO
RenderX

between the dependent variable—the rate of weight change—and various independent variables. Model 1 included age, gender, BMI, proportion of days with weight recording, and communication volume as the independent variables. Model 2 expanded upon Model 1 by introducing an interaction term between proportion of days with weight recording and chat count to examine their combined effect on weight change. The fit of Models 1 and 2 was compared to assess the contribution of the interaction term in explaining the variability in weight change. Upon identifying a significant interaction effect, a simple slope analysis was performed to further investigate the nature of this interaction. The proportion of days with weight recording was categorized into 3 groups based on SD, which included values more than one SD above the mean (+1 SD group), the mean itself (mean group), and one SD below the mean (−1 SD group). All statistical analyses were performed using R (version 4.1.0; R Foundation for Statistical Computing).

Results

Baseline Characteristics of the Study Participants

A total of 3999 participants were assessed for eligibility, and 90 dropped out during the program. Thus, 3909 participants (97.7%) completed the 6-month programme. Of these, 1486 were excluded from the analysis due to missing data, including those who did not provide data at the end of the study. Consequently, the final analysis was restricted to 2423 participants who met the criteria for the complete-data analysis. Table 1 shows the characteristics of the 2423 participants included in this study. Baseline characteristics were compared between completers (n=2423) and excluded participants (n=1486). Participants with complete data were older, more likely to be male, and had a higher prevalence of hypertension, while excluded participants had a higher prevalence of diabetes mellitus (P<.01 for all). Participants with complete data also had lower BMI, lower HbA_{1c}, and greater app engagement (higher chat volume and proportion of days with weight recording; P<.01). Detailed results were provided in Multimedia Appendix 1.

Table . Participants' characteristics. Values are shown as mean (SD) for ordinal variables and counts (%) for categorical variables.

Participants' characteristics	Results (n=2423)
Age (years), mean (SD)	54.9 (6.8)
Gender, n (%)	
Men	2144 (88.5)
Women	279 (11.5)
NCDs[a]	
Hypertension, n (%)	1842 (76)
Diabetes mellitus, n (%)	1199 (49.5)
Dyslipidemia, n (%)	1510 (62.3)
Previous stroke, n (%)	79 (3.3)
Previous ischemic heart disease, n (%)	151 (6.2)
Condition at the start of the program, mean (SD)	
Body weight (kg)	83.4 (12.7)
BMI (kg/m^2)	29.2 (3.7)
Systolic blood pressure (mmHg)	130.9 (12.2)
Diastolic blood pressure (mmHg)	84.5 (9.4)
HbA_{1c} (%)	6.5 (1.1)
HDL[b] cholesterol (mg/dL)	51.5 (11.9)
LDL[c] cholesterol (mg/dL)	121.8 (30.3)
App-related factor, mean (SD)	
Chat volume (chats per week)	1.9 (1.8)
Proportion of days with weight recording (%)	78.6 (26.7)

[a]NCD: noncommunicable disease.

[b]HDL: high-density lipoprotein.

[c]LDL: low-density lipoprotein.

XSL•FO
RenderX

Factors Associated With the Rate of Weight Change

Hierarchical multiple regression analyses were conducted to determine the factors associated with the rate of weight change among the participants. In Model 1, the independent variables included age, gender, baseline BMI, proportion of days with weight recording, and communication volume. The analysis identified significant negative coefficients for baseline BMI (β=−.097; *P*<.001), proportion of days with weight recording(β=−.017; *P*<.001), and communication volume (β=−.193; *P*<.001). Model 2 extended Model 1 by incorporating an interaction term between proportion of days with weight recording and chat frequency, which assesses their combined effect on weight change. The results of Model 2 showed a significant negative interaction effect (β=−.006; *P*<.001), suggesting that higher proportion of days with weight recording combined with frequent chats substantially enhance weight loss. Furthermore, Models 1 and 2 have shown that the addition of the interaction term significantly enhanced model performance (Δ*F*=13.97; *P*<.001). Table 2 has more details.

Table . Results of hierarchical linear regression analysis for weight change.

	Model 1[a] (Adjusted R^2=0.06)		Model 2[b] (Adjusted R^2=0.06)	
	β	*P* value	β	*P* value
Age	.016	.07	.02	.06
Gender (reference: Men)	.398	.02	.39	.03
BMI	−.097	<.001	−.10	<.001
Proportion of days with weight recording	−.017	<.001	−.02	<.001
Volume of chat communication	−.193	<.001	−.15	<.001
Proportion of days with weight recording × volume of chat communication (interaction)	—[c]	—	−.01	<.001

[a]Model 1 includes age, gender, BMI, proportion of days with weight recording, and volume of chat communication as independent variables.

[b]Model 2 extends model 1 by adding an interaction term between proportion of days with weight recording and volume of chat communication.

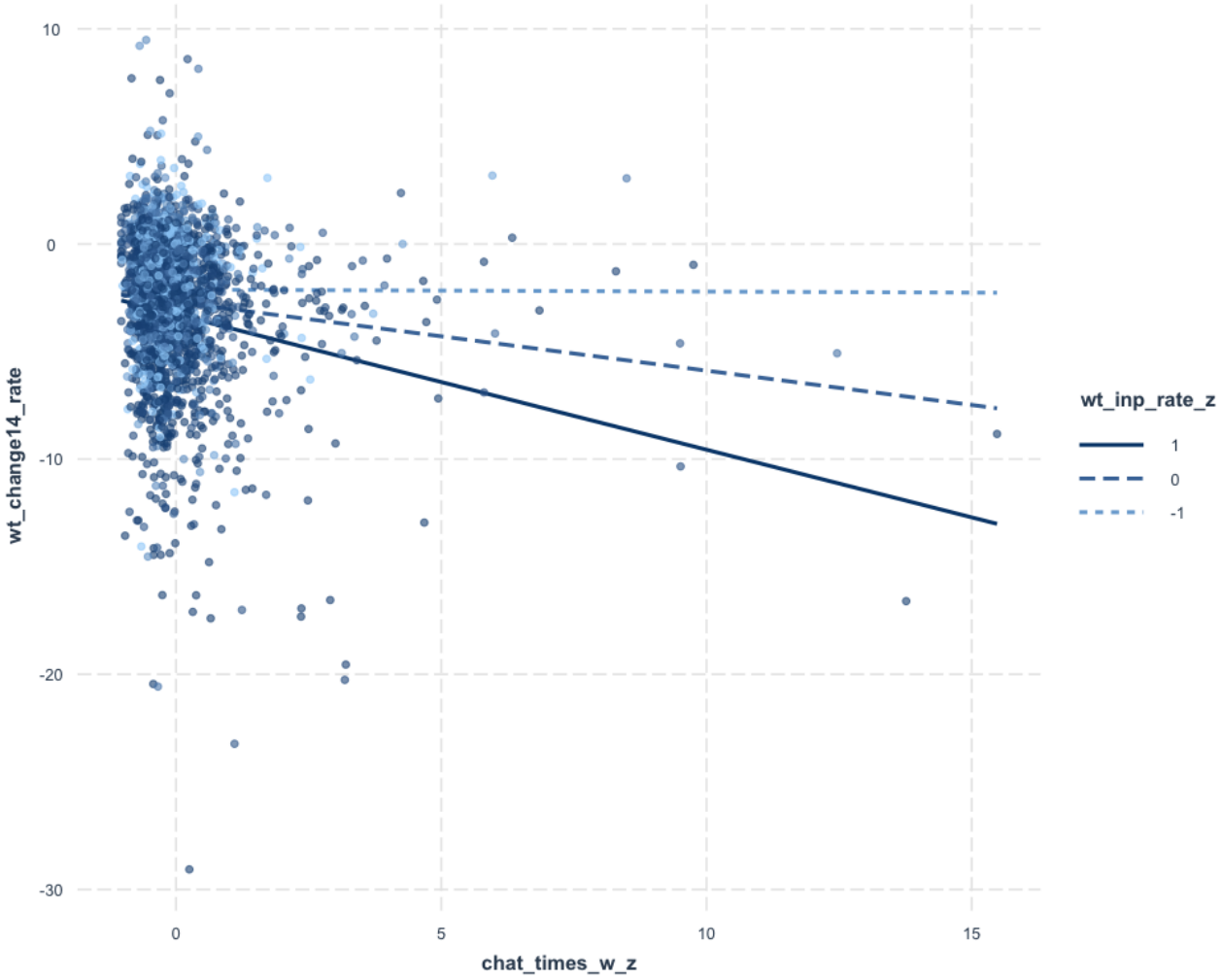
[c]Not applicable.

Simple Slopes Analysis

When the proportion of days with weight recording is −1 SD below the mean, the slope of weekly chat times on weight change rate is −0.01; this result is not statistically significant (*P*=.94). When the proportion of days with weight recording is at the mean (0 SD), the slope of weekly chat times on weight change rate is −0.32; this result is significant (*P*<.001). When the proportion of days with weight recording is +1 SD above the mean, the slope of weekly chat times on weight change rate is −0.6; this result is significant (*P*<.001) (see Figure 1).

XSL•FO
RenderX

Figure 1. The figure illustrates the interaction between proportion of days with weight recording (wt_inp_rate_z) and communication volume (chat_times_w_z) on weight change rate (wt_change14_rate).



Discussion

Principal Findings

This study aims to determine the factors associated with weight change rates in mHealth and assess the interaction of proportion of days with weight recording and communication volume on weight change rates. As a result, the regression analysis indicated that chat communication volume had a stronger association with weight change rates compared with the proportion of days with weight recording. Furthermore, the higher chat communication volume was associated with greater weight reduction in the case of high proportion of days with weight recording, whereas it was not significant when the proportion of days with weight recording were low.

The primary findings of this study were that the volume of chat communication and proportion of days with weight recording were associated with the weight change rate and that a significant interactive effect on weight loss was observed during the lifestyle modification program. Chat messages included both generic reminders, such as prompts to log weight, and personalized feedback based on participant progress and health data, as reflected by the high proportion of days with weight recording and the regular chat-based communication (1.9 exchanges per week) throughout the program. These interactions, combining standardized guidance and tailored support, may have helped participants stay motivated and accountable. In this study, the average proportion of days with weight recording over 6 months was 78.6% (1902/2423), indicating a higher user engagement. The lifestyle modification program in this study required biweekly telephone consultations based on the lifestyle data including weight and provided personalized feedback. Furthermore, this interactive approach likely contributed to encouraging the participants' proportion of days with weight recording. Previous studies have reported on the relationship between each indicator, proportion of days with weight recording, and communication volume, as well as the achievement of weight loss [11,12,14,15]. Weight measurement has been considered as one of the key indicators reflecting the user's engagement, which is related to weight loss outcomes in a mobile app intervention [12]. A study suggested that regular self-weighing patterns, particularly high and consistent self-weighing, are associated with greater weight loss and maintenance, as well as better adherence to energy intake and step goals in a 12-month behavioral weight loss intervention [13]. Another study of physical activity promotion determining the effectiveness of behavior modification support through gamified mobile apps has found that users with high engagement

XSL•FO
RenderX

showed greater improvements in physical activity [23]. However, it is still unclear how the interaction between the proportion of days with weight recording and the amount of communication affects weight loss due to lifestyle modifications. In this study, the interaction between the amount of communication and the effect of weight loss was higher when the proportion of days with weight recording was high, whereas no significant relation was observed between the amount of communication and the effect of weight loss when the proportion of days with weight recording was low. These findings suggest that frequent weight recording tends to increase the impact of communication with health care professionals, making personalized feedback more effective and contributing to greater weight loss.

Mainly 2 potential reasons could explain the findings of this study. First, a high proportion of days with weight recordings likely result in a richer and more continuous stream of data, allowing health care professionals to better understand the participants' condition and provide more tailored and precise advice. Conversely, a low proportion of days with weight recording tend to provide fragmented data, possibly resulting in more general feedback that may not be as effective. A systematic review by Berry et al [24] has reported that tailored advice in digital self-monitoring of physical activities and diet is an effective intervention for promoting weight loss in adults with obesity or overweight. Another study has reported that individualized text messages and treatment information provision tend to facilitate user behavioral change and the development of healthy behaviors [25]. The assumption from this study is that if health care providers can obtain large amounts of self-monitoring of health behaviors data from participants, it will be possible to provide highly personalized health plans and advice. This concept is in line with the studies reporting that lifestyle modification was promoted more by individualized feedback, reminders, and messages compared with standardized communication [26,27]. If only limited self-monitoring of health behaviors data were available, health care professionals would be unable to provide personalized advice, and the advice given would predominantly be general. Furthermore, a second possible explanation is the theory that frequent weight monitoring will foster the habit of self-monitoring, which will motivate participants to be more consistent in achieving their weight loss goals, thereby maximizing the effectiveness of communication with their instructors. A study by Conroy et al [28] investigated the relation between physical activity self-monitoring and weight loss and found that frequent self-monitoring led to increased physical activity and weight loss. Another study has found that users who stopped tracking their weight experienced weight gain and a decline in both their physical activity and activity tracking frequency [29]. These results suggest the association between weight monitoring and the motivation to achieve lifestyle modifications. Furthermore, patient–provider communication has been reported to affect adherence through trust and motivation [30]. This heightened motivation enhances the impact of communication with health care professionals as participants are more likely to heed the advice seriously and act upon it.

The findings emphasize the critical role of self-monitoring in the context of mHealth interventions. As digital health technologies become increasingly integrated into daily life, understanding the mechanisms and effects of self-monitoring on behavioral change is important. Self-monitoring is acknowledged as a cornerstone of behavioral change including lifestyle modification [31], involves individuals in self-care management, and enhances self-feedback through learning the link between daily lifestyle changes and health outcomes. The effectiveness of weight and lifestyle monitoring for weight reduction has been widely accepted even before the development of mHealth [24]. This study on mHealth intervention has confirmed the potential role of self-monitoring in the relation between mobile app-based communication and lifestyle modification, emphasizing the unique role of self-monitoring in the mHealth era. Thus, lifestyle monitoring using mobile apps may facilitate the positive cycle of behaviors and outcomes caused by personalized communication, which is a determinant of the effectiveness of lifestyle modifications in mHealth [24,32]. However, the high proportion of days with weight recording in this study may indicate the high readiness of the users for the program or their intrinsic motivation for lifestyle modification. Furthermore, the potential burden on the participants to consistently record their weight in daily living should also be considered to monitor long-term adherence. Therefore, further study is needed to investigate the causal relation between self-monitoring passively promoted by mHealth intervention, improved health care communication with health professionals, and subsequent behavioral change. In addition, future study should explore how chat interactions influence participant retention, particularly among those who do not experience early weight loss. Understanding whether chat engagement can help maintain motivation in these individuals could provide further findings for optimising mHealth interventions for long-term adherence.

Limitations

This study performed a retrospective analysis targeting participants in health programs provided by health insurance unions. Consequently, the participants were predominantly those who voluntarily joined the programs, the majority being males. This demographic imbalance may limit the generalizability of the findings. Future study should investigate whether similar trends are observed in more diverse populations, including elderly individuals and females. In addition, our analysis was conducted using a completers-only approach, meaning that participants who dropped out before completing the program were not included in the final analysis. Since previous study has shown that individuals who drop out of weight loss interventions tend to have poorer outcomes, our findings may not fully reflect the weight loss trajectories of all participants [33]. In fact, baseline characteristics differed significantly between participants with complete data and those with missing data. These differences suggest that participants with complete data may have been more intrinsically motivated or in better health, potentially influencing weight loss outcomes. Furthermore, this study measured the communication volume solely based on the number of chat messages sent and received, without considering the length or content of the messages. While

XSL•FO
RenderX

message frequency is a useful measure of engagement, we did not differentiate between participant-generated and provider-generated messages. Participant messages may indicate engagement and self-motivation, while provider messages likely focus on feedback and support. This lack of differentiation limits the ability to fully interpret the specific roles and contributions of each message type in the observed outcomes. Additional factors such as the duration of interactions, the frequency of health-related queries, and the number of proactive messages sent can provide further context into how communication affects weight loss outcomes. This limitation may affect the accuracy of assessing the impact of communication on weight change rate. Thus, future studies should develop and use more comprehensive methods for evaluating communication, including qualitative assessments of message content and context. Furthermore, determining the effectiveness of different communication strategies, such as personalized feedback and motivational interviewing techniques, could provide insights into optimizing engagement and outcomes. Another limitation is that the proportion of days with weight recording in this study was calculated as the proportion of days on which participants recorded their weight over the entire 6-month period. While this measure provides an overall indicator of engagement, it does not capture temporal variations in engagement patterns, such as whether participants maintained consistent recording behavior throughout the intervention or showed high initial engagement followed by decline. This distinction is important because sustained engagement is often a critical factor for achieving successful behavior modification in mHealth interventions. Future study should explore longitudinal patterns of weight recording behavior to provide a more nuanced understanding of compliance and engagement trends in mHealth-based weight management programs.

Conclusions

The results of this study have revealed an association between weight change rate and the interaction of proportion of days with weight recording and communication volume between health care staff and participants in a lifestyle modification program. When the proportion of days with weight recording was high, the relation between communication volume and weight change rate was strong. Conversely, when the proportion of days with weight recording was low, the relation between communication volume and weight change rate was weak. These findings suggested that the acquisition of self-monitoring of health behaviors and lifelog data was crucial for maximizing the effectiveness of the health care professional's support. Furthermore, these insights would be beneficial in the development and process management of lifestyle modification programs using mHealth.

Acknowledgments

The authors thank the participants of the mobile app-based program for their contributions. Special thanks to the study team at PREVENT Inc. for their support in data collection and management. This work was supported by JSPS KAKENHI (grant 23K16769) and the Taiju Life Social Welfare Foundation.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contributions

YH conceptualized the study, supervised data analysis, and drafted the manuscript. KS performed statistical analysis and contributed to manuscript preparation. SI performed statistical analysis and contributed to manuscript preparation. TA provided expertise on study design and critically revised the manuscript. MK assisted in data collection, manuscript writing, and revisions. TM provided oversight on methodology and reviewed the final manuscript.

Conflicts of Interest

This study was conducted as part of PREVENT, Inc's internal research activities. No external funding was received for this work. PREVENT, Inc provided access to anonymized program data and other necessary resources for the research. The company had no influence on the study design, data analysis, interpretation of findings, or the preparation of the manuscript, which were conducted independently by the authors.

Multimedia Appendix 1
Comparison between participants with complete data and excluded participants.
[DOCX File, 19 KB - i-jmr_v14i1e65863_app1.docx]

References

1. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222. [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]

XSL•FO
RenderX

2. Okunogbe A, Nugent R, Spencer G, Powis J, Ralston J, Wilding J. Economic impacts of overweight and obesity: current and future estimates for 161 countries. BMJ Glob Health 2022 Sep;7(9):e009773. [doi: 10.1136/bmjgh-2022-009773] [Medline: 36130777]
3. Kotseva K, De Backer G, De Bacquer D, et al. Primary prevention efforts are poorly developed in people at high cardiovascular risk: a report from the European Society of Cardiology EURObservational Research Programme EUROASPIRE V survey in 16 European countries. Eur J Prev Cardiol 2021 May 8;28(4):370-379. [doi: 10.1177/2047487320908698] [Medline: 33966079]
4. Powers WJ, Rabinstein AA, Ackerson T, et al. Guidelines for the early management of patients with acute ischemic stroke: 2019 update to the 2018 guidelines for the early management of acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. Stroke 2019 Dec;50(12):e344-e418. [doi: 10.1161/STR.0000000000000211] [Medline: 31662037]
5. Bairey Merz CN, Alberts MJ, Balady GJ, et al. ACCF/AHA/ACP 2009 competence and training statement: a curriculum on prevention of cardiovascular disease: a report of the American College of Cardiology Foundation/American Heart Association/American College of Physicians Task Force on Competence and Training (Writing Committee to Develop a Competence and Training Statement on Prevention of Cardiovascular Disease): developed in collaboration with the American Academy of Neurology; American Association of Cardiovascular and Pulmonary Rehabilitation; American College of Preventive Medicine; American College of Sports Medicine; American Diabetes Association; American Society of Hypertension; Association of Black Cardiologists; Centers for Disease Control and Prevention; National Heart, Lung, and Blood Institute; National Lipid Association; and Preventive Cardiovascular Nurses Association. Circulation 2009 Sep 29;120(13):e100-e126. [doi: 10.1161/CIRCULATIONAHA.109.192640] [Medline: 19770387]
6. Pearson GJ, Thanassoulis G, Anderson TJ, et al. 2021 Canadian Cardiovascular Society guidelines for the management of dyslipidemia for the prevention of cardiovascular disease in adults. Can J Cardiol 2021 Aug;37(8):1129-1150. [doi: 10.1016/j.cjca.2021.03.016] [Medline: 33781847]
7. Zomer E, Gurusamy K, Leach R, et al. Interventions that cause weight loss and the impact on cardiovascular risk factors: a systematic review and meta-analysis. Obes Rev 2016 Oct;17(10):1001-1011. [doi: 10.1111/obr.12433] [Medline: 27324830]
8. Dambha-Miller H, Day AJ, Strelitz J, Irving G, Griffin SJ. Behaviour change, weight loss and remission of Type 2 diabetes: a community-based prospective cohort study. Diabet Med 2020 Apr;37(4):681-688. [doi: 10.1111/dme.14122] [Medline: 31479535]
9. Jung J, Cho I. Promoting physical activity and weight loss with mhealth interventions among workers: systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2022 Jan 21;10(1):e30682. [doi: 10.2196/30682] [Medline: 35060913]
10. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 4;20(5):e162. [doi: 10.2196/jmir.9751] [Medline: 29728346]
11. Patel ML, Wakayama LN, Bennett GG. Self-monitoring via digital health in weight loss interventions: a systematic review among adults with overweight or obesity. Obesity (Silver Spring) 2021 Mar;29(3):478-499. [doi: 10.1002/oby.23088] [Medline: 33624440]
12. Carey A, Yang Q, DeLuca L, Toro-Ramos T, Kim Y, Michaelides A. The relationship between weight loss outcomes and engagement in a mobile behavioral change intervention: retrospective analysis. JMIR Mhealth Uhealth 2021 Nov 8;9(11):e30622. [doi: 10.2196/30622] [Medline: 34747706]
13. Zheng Y, Burke LE, Danford CA, Ewing LJ, Terry MA, Sereika SM. Patterns of self-weighing behavior and weight change in a weight loss trial. Int J Obes 2016 Sep;40(9):1392-1396. [doi: 10.1038/ijo.2016.68]
14. 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]
15. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. Digit Health 2020;6:2055207620914427. [doi: 10.1177/2055207620914427] [Medline: 32269830]
16. Zhaoyang R, Sliwinski MJ, Martire LM, Smyth JM. Social interactions and physical symptoms in daily life: quality matters for older adults, quantity matters for younger adults. Psychol Health 2019 Jul;34(7):867-885. [doi: 10.1080/08870446.2019.1579908] [Medline: 30821176]
17. Godino JG, Merchant G, Norman GJ, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diabetes Endocrinol 2016 Sep;4(9):747-755. [doi: 10.1016/S2213-8587(16)30105-X] [Medline: 27426247]
18. Little P, Stuart B, Hobbs FR, et al. Randomised controlled trial and economic analysis of an internet-based weight management programme: POWeR+ (Positive Online Weight Reduction). Health Technol Assess 2017 Jan;21(4):1-62. [doi: 10.3310/hta21040] [Medline: 28122658]
19. Arnold J, Bailey CP, Evans WD, Napolitano MA. Application of Mcguire's model to weight management messages: measuring persuasion of facebook posts in the healthy body, healthy u trial for young adults attending university in the

XSL•FO
RenderX

United States. Int J Environ Res Public Health 2022 Nov 1;19(21):14275. [doi: 10.3390/ijerph192114275] [Medline: 36361162]
20. Young AS, Cohen AN, Goldberg R, et al. Improving weight in people with serious mental illness: the effectiveness of computerized services with peer coaches. J Gen Intern Med 2017 Apr;32(Suppl 1):48-55. [doi: 10.1007/s11606-016-3963-0] [Medline: 28271427]
21. Kanai M, Toda T, Yamamoto K, Akimoto M, Hagiwara Y. A mobile health-based disease management program improves blood pressure in people with multiple lifestyle-related diseases at risk of developing vascular disease - a retrospective observational study. Circ Rep 2022 Jul 8;4(7):322-329. [doi: 10.1253/circrep.CR-22-0024] [Medline: 35860354]
22. Muramoto A, Matsushita M, Kato A, et al. Three percent weight reduction is the minimum requirement to improve health hazards in obese and overweight people in Japan. Obes Res Clin Pract 2014;8(5):e466-e475. [doi: 10.1016/j.orcp.2013.10.003] [Medline: 25263836]
23. Edney S, Ryan JC, Olds T, et al. User engagement and attrition in an app-based physical activity intervention: secondary analysis of a randomized controlled trial. J Med Internet Res 2019 Nov 27;21(11):e14645. [doi: 10.2196/14645] [Medline: 31774402]
24. Berry R, Kassavou A, Sutton S. Does self-monitoring diet and physical activity behaviors using digital technology support adults with obesity or overweight to lose weight? A systematic literature review with meta-analysis. Obes Rev 2021 Oct;22(10):e13306. [doi: 10.1111/obr.13306] [Medline: 34192411]
25. Gonzalez C, Early J, Gordon-Dseagu V, Mata T, Nieto C. Promoting culturally tailored mHealth: a scoping review of mobile health interventions in latinx communities. J Immigr Minor Health 2021 Oct;23(5):1065-1077. [doi: 10.1007/s10903-021-01209-4] [Medline: 33988789]
26. Hutchesson MJ, Tan CY, Morgan P, Callister R, Collins C. Enhancement of self-monitoring in a web-based weight loss program by extra individualized feedback and reminders: randomized trial. J Med Internet Res 2016 Apr 12;18(4):e82. [doi: 10.2196/jmir.4100] [Medline: 27072817]
27. Lewis E, Hassmén P, Pumpa KL. Participant perspectives of a telehealth trial investigating the use of telephone and text message support in obesity management: a qualitative evaluation. BMC Health Serv Res 2021 Jul 9;21(1):675. [doi: 10.1186/s12913-021-06689-6] [Medline: 34243772]
28. Conroy MB, Yang K, Elci OU, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574. [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]
29. Frie K, Hartmann-Boyce J, Jebb S, Oke J, Aveyard P. Patterns in weight and physical activity tracking data preceding a stop in weight monitoring: observational analysis. J Med Internet Res 2020 Mar 17;22(3):e15790. [doi: 10.2196/15790] [Medline: 32181749]
30. Young HN, Len-Rios ME, Brown R, Moreno MM, Cox E. How does patient-provider communication influence adherence to asthma medications? Patient Educ Couns 2017 Apr;100(4):696-702. [doi: 10.1016/j.pec.2016.11.022] [Medline: 27916463]
31. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: a report of the American College of Cardiology/American Heart Association Task Force on clinical practice guidelines. Circulation 2019 Sep 10;140(11):e596-e646. [doi: 10.1161/CIR.0000000000000678] [Medline: 30879355]
32. Lahtio H, Rintala A, Immonen J, Sjögren T. The effectiveness of physical activity-promoting web- and mobile-based distance weight loss interventions on body composition in rehabilitation settings: systematic review, meta-analysis, and meta-regression analysis. J Med Internet Res 2022 Mar 24;24(3):e25906. [doi: 10.2196/25906] [Medline: 35323126]
33. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011 Nov;12(11):912-934. [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]

Abbreviations

mHealth: mobile health
NCD: noncommunicable disease

Edited by TDA Cardoso; submitted 28.08.24; peer-reviewed by G Jerome, M Yang; revised version received 15.02.25; accepted 18.02.25; published 27.03.25.

Please cite as:
Hagiwara Y, Adachi T, Kanai M, Shimizu K, Ishida S, Miki T
Interactive Effects of Weight Recording Frequency and the Volume of Chat Communication With Health Care Professionals on Weight Loss in mHealth Interventions for Noncommunicable Diseases: Retrospective Observational Study
Interact J Med Res 2025;14:e65863
URL: https://www.i-jmr.org/2025/1/e65863
doi:10.2196/65863

XSL•FO
RenderX

© Yuta Hagiwara, Takuji Adachi, Masashi Kanai, Kotoe Shimizu, Shinpei Ishida, Takahiro Miki. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.3.2025. 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

Application of Machine Learning and Emerging Health Technologies in the Uptake of HIV Testing: Bibliometric Analysis of Studies Published From 2000 to 2024

Musa Jaiteh[1*], MSc; Edith Phalane[1*], PhD; Yegnanew A Shiferaw[2*], PhD; Lateef Babatunde Amusa[3,4*], PhD; Hossana Twinomurinzi[3*], PhD; Refilwe Nancy Phaswana-Mafuya[1*], PhD

[1]South African Medical Research Council/University of Johannesburg Pan African Centre for Epidemics Research Extramural Unit, Faculty of Health Sciences, University of Johannesburg, Auckland Park Bunting Road Campus, PO Box 524, Auckland Park, Johannesburg, South Africa

[2]Department of Statistics, Faculty of Science, University of Johannesburg, Johannesburg, South Africa

[3]Center of Applied Data Science, University of Johannesburg, Johannesburg, South Africa

[4]Department of Statistics, University of Ilorin, Ilorin, Nigeria

[*]all authors contributed equally

Corresponding Author:
Refilwe Nancy Phaswana-Mafuya, PhD
South African Medical Research Council/University of Johannesburg Pan African Centre for Epidemics Research Extramural Unit, Faculty of Health Sciences, University of Johannesburg, Auckland Park Bunting Road Campus, PO Box 524, Auckland Park, Johannesburg, South Africa

Abstract

Background: The global targets for HIV testing for achieving the Joint United Nations Programme on HIV/AIDS (UNAIDS) 95-95-95 targets are still short. Identifying gaps and opportunities for HIV testing uptake is crucial in fast-tracking the second (initiate people living with HIV on antiretroviral therapy) and third (viral suppression) UNAIDS goals. Machine learning and health technologies can precisely predict high-risk individuals and facilitate more effective and efficient HIV testing methods. Despite this advancement, there exists a research gap regarding the extent to which such technologies are integrated into HIV testing strategies worldwide.

Objective: The study aimed to examine the characteristics, citation patterns, and contents of published studies applying machine learning and emerging health technologies in HIV testing from 2000 to 2024.

Methods: This bibliometric analysis identified relevant studies using machine learning and emerging health technologies in HIV testing from the Web of Science database using synonymous keywords. The Bibliometrix R package was used to analyze the characteristics, citation patterns, and contents of 266 articles. The VOSviewer software was used to conduct network visualization. The analysis focused on the yearly growth rate, citation analysis, keywords, institutions, countries, authorship, and collaboration patterns. Key themes and topics were driven by the authors' most frequent keywords, which aided the content analysis.

Results: The analysis revealed a scientific annual growth rate of 15.68%, with an international coauthorship of 8.22% and an average citation count of 17.47 per document. The most relevant sources were from high-impact journals such as the *Journal of Internet Medicine Research*, *JMIR mHealth and uHealth*, *JMIR Research Protocols*, *mHealth*, *AIDS Care-Psychological and Socio-Medical Aspects of AI*, and *BMC Public Health*, and *PLOS One*. The United States of America, China, South Africa, the United Kingdom, and Australia produced the highest number of contributions. Collaboration analysis showed significant networks among universities in high-income countries, including the University of North Carolina, Emory University, the University of Michigan, San Diego State University, the University of Pennsylvania, and the London School of Hygiene and Tropical Medicine. The discrepancy highlights missed opportunities in strategic partnerships between high-income and low-income countries. The results further demonstrate that machine learning and health technologies enhance the effective and efficient implementation of innovative HIV testing methods, including HIV self-testing among priority populations.

Conclusions: This study identifies trends and hotspots of machine learning and health technology research in relation to HIV testing across various countries, institutions, journals, and authors. The trends are higher in high-income countries with a greater focus on technology applications for HIV self-testing among young people and priority populations. These insights will inform future researchers about the dynamics of research outputs and help them make scholarly decisions to address research gaps in this field.

(Interact J Med Res 2025;14:e64829) doi:10.2196/64829

XSL•FO
RenderX

KEYWORDS

bibliometric analysis; machine learning; mHealth; technology; HIV prevention; HIV testing; application; bibliometric; HIV; epidemiology; AIDS; self-testing; sexually transmitted disease; quantitative; bibliography; content analysis; mobile health

Introduction

HIV remains a global public health burden, affecting more than 39 million people living with HIV, with approximately 60% of people living with HIV living in sub-Saharan Africa (SSA) [1,2]. HIV/AIDS has claimed 42.3 million lives since the beginning of the epidemic and 630,000 lives in 2023 alone [1]. In addition, the continuous transmission and heightened rates of new HIV infections across different countries remain a significant concern [1,2]. There have been ongoing efforts to curb HIV as a public health threat, among which is the Joint United Nations Programme on HIV/AIDS (UNAIDS) goal of ending AIDS by 2030 [3]. UNAIDS and the World Health Organization (WHO) highlight the significance of prioritizing HIV testing for the realization of this goal [3,4]. Getting tested is the only way to identify people living with HIV for early initiation of antiretroviral therapy to reduce HIV transmission rates [3,4].

Current trends indicate a significant gap in the uptake of HIV testing globally, as only 86% of people living with HIV knew their HIV status by the end of 2023 [2]. This leaves approximately 5.4 million people living with HIV untested, which falls short of the UNAIDS interim target of testing 95% of people living with HIV by 2025 [2,3]. Conventional HIV testing and prevention programs have shown progress, but they fail to address stigma and social discrimination among key populations (KPs) [2-5]. Due to this challenge, HIV testing rates remain low among KPs, who bear the highest burden of HIV incidence [2,5]. In response to this issue, the WHO released updated comprehensive guidelines in 2016 to improve HIV testing coverage while using new strategies, including HIV self-testing (HIVST) [5]. The guidelines are continuously updated to integrate emerging technologies, providing more convenient options for HIV testing and helping to fast-track the UNAIDS 2030 goal.

The emergence of technology-driven health care interventions in recent years has resulted in remarkable advancements, especially in the realm of HIV testing [6]. Among these innovations, machine learning emerged as a noteworthy tool for enhancing the efficacy, accuracy, and accessibility of HIV testing services [7-10]. Machine learning involves the use and development of computer systems that can learn and adapt without explicit instructions by using algorithms to analyze and draw conclusions from data patterns [10].

Machine learning can improve HIV testing in a variety of ways. This methodology accurately predicts latent factors influencing HIV testing by analyzing complex data that could not be explored by the common approaches. Precisely, machine learning identifies individuals at the highest risk of HIV for tailored and cost-effective HIV testing programs to optimize resource allocation [7-11]. This is extremely beneficial, especially for low-income countries with limited resources to combat high HIV rates and other prevailing health issues. A study by Xu et al [11] confirmed that machine learning models such as logistic regression, support vector machines, random forest (RF), Extreme Gradient Boosting (XGBoost), elastic net, and k-nearest neighbor, among other algorithms, were reliable in predicting future HIV risk using complex datasets. Similarly, a retrospective analysis conducted in the United States [12] discovered that RF models seem promising for identifying important predictors of HIV testing uptake. The model's integration into future HIV prevention and treatment research, as well as intervention program evaluations, was highly encouraged [12]. Datasets from 4 SSA countries were analyzed by XGBoost, support vector machines, RF, elastic net, k-nearest neighbor, and light gradient boosting machine algorithms to accurately identify high-priority groups susceptible to HIV for enhancing targeted HIV testing [7].

Recently, more advanced machine learning technologies have emerged as a game changer for HIV testing concerns such as privacy and accessibility issues with high rates of false positivity/negativity. For instance, the integration of HIVST into machine learning, digital health, and mobile devices is gaining popularity, especially among KPs who usually avoid public services due to stigma [13]. Consistently, the WHO recommends HIVST as a convenient and discrete choice for HIV testing, citing its dependability and excellent performance as key factors for its success [5]. In 2017, van Heerden et al [14] installed artificial intelligence (AI) conversational agents into Android devices to facilitate HIVST and counseling. It was interesting to find out that the South Africans who participated in the study felt the device was more convenient for HIV counseling and testing than the usual procedure. Mobile health (mHealth), which involves using mobile devices to enhance health care services and research, has proven advantageous when incorporated with machine learning [15]. A machine learning mHealth technique was used to detect false-positive HIV test results in another South African study [15]. The technology demonstrated high levels of sensitivity (97.8%) and specificity (100%) compared to conventional visual interpretations of HIV rapid diagnostic tests [15]. Similar advancements were made by [16,17] with their HealthPulse AI technologies, which outperformed human interpretation of HIV test results and provided an opportunity to reduce mental health disturbances associated with false-positive test results.

Leveraging machine learning methods and health technologies can improve HIV testing outcomes by prioritizing high-risk individuals and enhancing the accuracy of HIV testing devices. Despite these advancements, a research gap remains regarding the integration of machine learning techniques into HIV testing interventions. This bibliometric analysis aims to address this gap by examining trends in published studies, journal performances, and collaboration networks among researchers, institutions, and countries regarding using machine learning and emerging health technologies in HIV testing from 2000 to 2024. By identifying key patterns and gaps, the findings will inform future research directions and support more effective

XSL•FO
RenderX

implementation of machine learning and emerging health technologies in HIV testing.

Methods

Study Approach and Search Strategy

This bibliometric analysis was conducted to understand trends, patterns, and collaboration networks in published studies focused on using machine learning and emerging health technologies in HIV testing from 2000 to 2024. A search string was developed using synonyms of "machine learning" OR "health technologies" and "HIV testing" with Medical Subject Headings (MeSH) terms in PubMed and published keywords from related studies. The articles used for this bibliometric analysis were gathered from the Web of Science (WoS) database. The WoS is one of the world's top databases for scientific research and citation data, encompassing a wide range of areas such as medicine and health care [18-20]. It has been recognized as a legitimate platform for bibliometric analysis [19-23].

The search query was entered into the WoS search box and filtered by topics, which captured a total of 708 articles at the initial search, and the article list was refined by the inclusion and exclusion criteria listed in Table 1. The study analyzed 266 articles that met the inclusion criteria.

Table . Inclusion and exclusion criteria.

	Inclusion criteria	Exclusion criteria
Document type	Original and review articles	Proceedings, early access, letters, meeting abstracts, letters, editorial materials, data papers, retracted materials, and corrections.
Years of publication	January 2000 to December 2024	Before January 2000
Language	English	Catalan, French, German, and Spanish
Context	Machine learning and emerging health technologies in HIV testing	Machine learning and emerging health technologies in pre-exposure prophylaxis, antiretroviral therapy, postexposure prophylaxis, voluntary medical male circumcision, HIV education, and unrelated topics.
Research area	Human medicines and health care–related subjects	Religion, government law, history, philosophy of science, arts, humanities, and other topics

The search string that was used to retrieve studies that used machine learning and emerging health technologies in HIV testing from the WoS is detailed here:

("HIV testing" OR "AIDS Testing" OR "HIV Diagnosis" OR "Rapid HIV testing" OR "HIV Self-Testing" OR "HIV Screening" OR "Human Immunodeficiency Virus Testing" OR "HIV counsel$ing and testing")

AND

("Machine Learning" OR "Artificial Intelligence" OR "Unsupervised Learning" OR "Supervised Learning" OR "Model* Algorithm*" OR "Deep Learning" OR "Neural Network*" OR "Support Vector Machine*" OR "Random Forest" OR "Decision Tree*" OR "Convolutional Neural Networks" OR "Recurrent Neural Networks" OR "artificial neural networks" OR "LASSO" OR "XGBoost" OR "ensemble learning" OR "reinforce* learning" OR "Naive Bayes*" OR "AI chatbot" OR "Large language model*" OR "principal component analysis" OR "natural language processing" OR "Health technology" OR "health informatics" OR "Bayesian network*" OR "mHealth" OR "Digital Health" OR "eHealth" OR "Video gam*" OR "Telehealth" OR "Telemedicine" OR "mobile app*")

Inclusion and Exclusion Criteria

The inclusion and exclusion criteria outlined in Table 1 were carefully selected to ensure a relevant and comprehensive dataset. Studies were considered for this bibliometric analysis if they applied machine learning and health technologies in HIV testing, diagnosis, or predictive models. Articles that used machine learning and emerging health technologies in other HIV interventions, such as treatment and education programs, were excluded. Moreover, the analysis focused on original and published studies between January 2000 and December 2024. The criteria used in this study were to align the focus of this analysis with the study objective.

Data Processing and Analysis

Data Cleaning and Screening

The bibliographic data were initially cleaned to remove duplicates and exclude studies that fell outside the scope of this study (refer to Table 1). Filters on the WoS database were used to remove articles published before 2000, studies conducted in languages other than English, and studies that were not original or reviewed. Of the 356 articles, 342 remained in the sample and were exported into the Covidence web software for screening. The Covidence software automatically removed 5 duplicates, and 2 independent reviewers (MJ and EP) conducted a comprehensive screening to retain eligible articles in the sample. Articles were screened by title and abstract, and 56 ineligible articles were removed. The full text of the remaining articles was also screened; of these, 266 were selected for the final sample for analysis. The two reviewers independently screened the articles at each stage and resolved disagreements through consensus. The screening process is presented in the Results section.

XSL•FO
RenderX

Data Analysis

This bibliometric analysis was performed using the Biliometrix R package (version 4.3.2; Posit Software, PBC) [24], while the VOSviewer software (version 1.6.20) was used for network visualization.

After the data were screened, cleaned, and refined according to the inclusion criteria, 266 references were exported from Covidence as CSV files and exported to Zotero (open-source bibliographic software developed by the global community) for conversion into retained and converted plain text (TXT) files, acceptable file format for the bibliometric software. Thereafter, the downloaded TXT file was uploaded into the Bibliometrix online software (Biblioshiny) [24], an R package (version 4.3.2; Posit Software, PBC). The cleaned bibliographic data was loaded into Biblioshiny to conduct performance analysis and science mapping, and VOSviewer aided in the network visualization.

Performance Analysis

Performance analysis examines the output and impact of research constituents by quantifying their citation and publication-related metrics [19,20]. It summarizes the characteristics of studies by assessing their productivity (number of publications), impact (citations), and contributors (authors, journals, countries, institutions, subject areas, etc) [19,20,25]. This study first provides a descriptive summary of the selected articles and then presents metrics such as the annual number of scientific productions, top-cited articles, and most productive authors, journals, affiliations, and countries in the field of machine learning in HIV testing.

Various metrics were used to measure the bibliographic sample's productivity, influence, and impact. For instance, h-index and m-index were used to evaluate the performance of influential authors and journals. According to previous studies [23,25,26], the h-index is referred to as the "highest number of h-index, such that the individual has published 'h' papers that each has been cited 'h' times." This means that if an author has published 20 articles and each is cited 12 times, the author's h-index is 12 [26]. The m-index is a variation of the h-index, showing the h-index per year since the author's first publication [23]. These two metrics are suitable for measuring the impact of authors and journals [23,26]; thus, they were calculated in this study using the Bibliometrix software. Additionally, the number of global citations for the articles was also used to examine their impact. We further assessed the performances of the authors' countries and affiliations using the frequency and number of publications on studies related to machine learning in HIV testing between 2000 and 2024.

Science Mapping

Science mapping, on the other hand, investigates the relationships between research constituents by analyzing the topic's conceptual, intellectual, and social structure [19,20]. We performed various analyses, including co-citation analysis of cited references, co-occurrence analysis of author keywords, and coauthorship analysis (international collaboration networks). These analyses explore the collaboration patterns between publications, citations, authors, institutions, and countries, as well as describe existing and future relations among topics [20]. These networks explain the research area's conceptual, intellectual, and social structures, forming the basis for developing key themes for this study. The VOSviewer software (version 1.6.20) provides a visualized and comprehensive understanding of the relationships between different networks within the dataset [20]. Normalization was applied at full counting for co-citation, co-occurrence, and international collaboration analyses. A maximum of 25 countries per document and a minimum of 5 documents per country were established as eligibility criteria, resulting in 53 eligible countries, of which 13 met the collaboration threshold. A total of 598 institutions appeared no more than 25 times in each document, with a minimum requirement of 10 documents per institution; ultimately, 16 institutions met this threshold. The minimum citation requirement for each cited reference was set at 15, and 11 of 8911 cited references satisfied this criterion. Additionally, with a minimum of 10 occurrences required per keyword, 19 of 700 keywords achieved the threshold for co-occurrence analysis. The research constituents, such as country, institution, keyword, author, and document source, were each represented by network nodes. The size of the nodes represents the occurrence frequencies, while the thickness of the links between the nodes determines the strength of the association between networks [23]. These analyses generated key concepts and thematic areas explored in the content analysis. Clustering was used to identify themes from author keywords and the names of the authors' institutions, as well as country names, within the same cluster in each network. The cluster size was set to 1 with a resolution of 1.00.

Content Analysis

The study explores the content of articles with the authors' most frequent keywords, which are also within the scope of the key themes generated by the keywords. The word cloud and co-occurrence analyses aided this process. By analyzing the authors' co-occurring keywords, we examined various clusters of relevant keywords that formed the basis of the content analysis. The content analysis aided in gaining deeper insights into the patterns and dynamics of the intersection between machine learning/health technologies and HIV testing, as well as suggested future directions from the identified gaps. A qualitative summary of the content analysis presented various topics, which can be found in the Results section.

Ethical Considerations

This study forms a part of a doctoral study by the first author (MJ) titled "Integration of machine learning algorithms to predict HIV testing associations using repeated cross-sectional survey data in an adult South African population: An HIV testing predictive model." The study was reviewed and approved by the University of Johannesburg (UJ) Research and Ethics Committee (REC; ethics approval number REC-2725 - 2024). Since no primary data were collected, the UJ REC gave a waiver for informed consent. Furthermore, the doctoral study falls under an umbrella project funded by the South African Medical Research Council (SAMRC) under the SAMRC/UJ Pan African Centre for Epidemics Research (PACER) Extramural Unit, titled "Harnessing big heterogeneous data to evaluate the

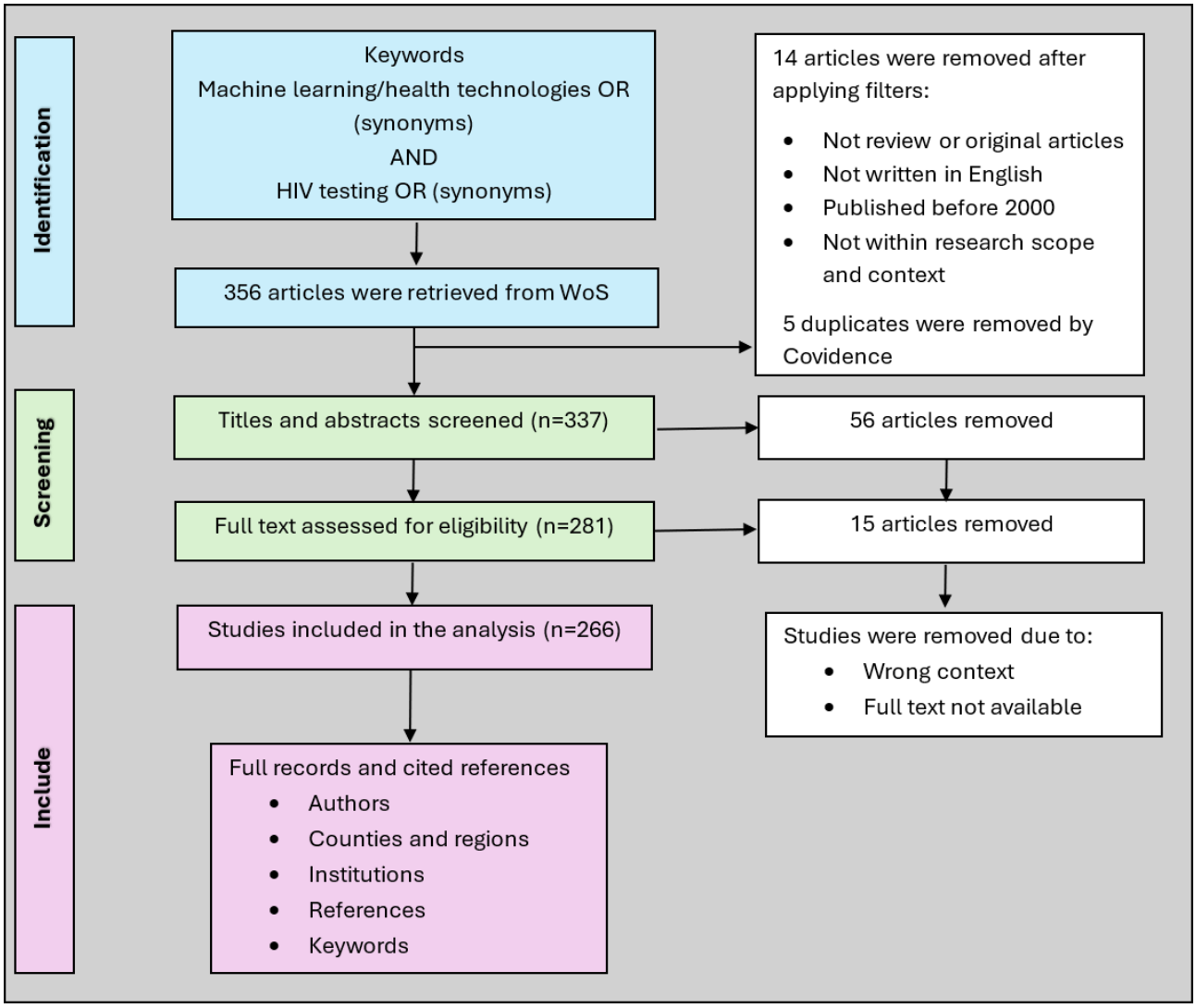
XSL•FO
RenderX

potential impact on HIV responses among the key populations in generalized epidemic settings in Sub-Saharan Africa" (ethics approval number REC-1504 - 2022).

Results

This study analyzed 266 articles following a comprehensive screening as shown in Figure 1.

Figure 1. Flowchart showing the search strategy and bibliometric analysis process. WoS: Web of Science.



Performance Analysis: Characteristics of the Study Sample

Table 2 summarizes the main characteristics of the articles included in this bibliometric analysis. The analyzed bibliographic dataset comprised 266 articles from 96 sources with 1658 authors and 8918 references. Most (91%) of the documents were original articles, while 9% were review articles. The results show a steady increase in the scientific production of studies using machine learning and emerging health technologies in HIV testing between 2000 and 2024, with an annual growth rate of 15.68%.

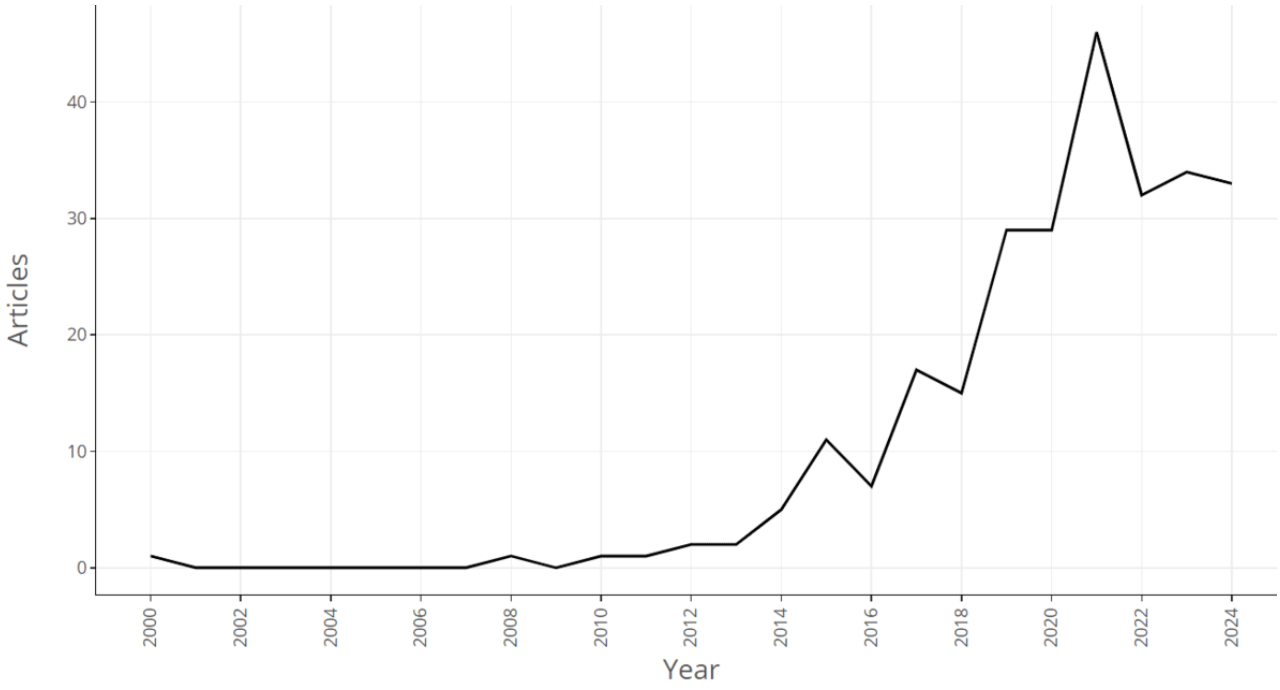
XSL•FO
RenderX

Table . Main information of the extracted bibliographic data of published articles on machine learning in HIV testing from 2000 to 2024.

Description	Results
Main information about data	
Time span (years)	2000-2024
Sources (journals, books, etc), n	96
Documents, n	266
Annual growth rate, %	15.68
Document average age (years)	3.85
Average citations per document, n	17.47
References, n	8918
Document contents	
Keywords Plus (generated by the indexing database), n	563
Authors' keywords, n	700
Authors	
Authors, n	1658
Authors of single-authored documents, n	2
Authors' collaboration	
Single-authored documents, n	2
Coauthors per document, n	8.22
International coauthorships, %	42.48
Document types	
Original articles, n	241
Reviews, n	25

Figure 2 highlights that annual scientific production in this field became more rapid between 2013 and 2021, but research outputs were comparatively low before 2013. The highest number of publications was recorded in 2021.

Figure 2. Annual scientific production of articles focused on the application of machine learning in HIV testing.



XSL•FO
RenderX

The output in Table 3 presents the top 10 most productive and influential journals that publish studies focused on machine learning and emerging health technologies in HIV testing–related topics. Among them, *AIDS and Behavior* ranked at the top of the list, marked by the highest h-index (10), number of citations (494), and number of publications (19). Other sources that are very productive and influential in this research area are *Journal of Medical Internet Research* (h-index=10), *JMIR mHealth and uHealth* (h-index=9), *JMIR Research Protocols* (h-index=8), *mHealth* (h-index 8), *AIDS Care-Psychological and Socio-Medical Aspects of AI* (h-index=6), and *BMC Public Health* (h-index=6), *PLOS One* (h-index=6), *AIDS Education and Prevention* (h-index=5), and *JMIR Formative Research* (h-index=5).

Table . Top 10 most productive publication sources by h-index.

Source	H-index	M-index	Total citations	Number of publications	Year of first included study
AIDS and Behavior	10	0.909	494	19	2014
Journal of Medical Internet Research	10	0.909	372	14	2014
JMIR mHealth and uHealth	9	0.818	310	14	2014
JMIR Research Protocols	8	1.333	264	18	2019
mHealth	8	1.6	138	10	2020
AIDS Care-Psychological and Socio-Medical Aspects of AI	6	0.6	120	7	2015
BMC Public Health	6	0.667	120	9	2016
PLOS One	6	0.4	160	9	2010
AIDS Education and Prevention	5	0.5	144	6	2015
JMIR Formative Research	5	1.25	51	12	2021

In Table 4, the top 10 most productive and influential authors were ranked by h-index, and their scores were based on studies related to machine learning and emerging health technologies in HIV testing. Stephenson has the highest h-index (10), followed by Bauermeister and Sullivan, who each scored 9. Additionally, Hightow-Weidman (h-index=8), Horvath (h-index=8), Bukusi (h-index=7), and Muessig (h-index=7) are also very influential in this field. Stephenson has a total of 421 citations with 16 publications of articles centered on machine learning (technology) in conjunction with HIV testing.

Table . Top 10 most productive and influential authors by h-index.

Author	H-index	M-index	Total citations	Number of publications	First publication year of an included study
Stephenson R	11	1	421	16	2014
Bauermeister JA	9	1.125	176	11	2017
Sullivan PS	9	0.818	406	16	2014
Hightow-Weidman L	8	0.1333	226	10	2019
Horvath KJ	8	0.889	163	13	2016
Bukusi EA	7	0.636	388	8	2014
Muessig KE	7	0.7	223	7	2015
Mustanski B	6	0.6	225	7	2015
Tucker JD	6	0.6	403	11	2015
Cohen CR	5	0.455	353	5	2014

In terms of the most impactful and influential articles, Table 5 presents the top 10 papers that have applied machine learning and emerging health technologies in HIV testing–related studies across the globe. A study by Schnall et al [27] published in the *Journal of Medical Internet Research* was the most influential paper with 129 total citations, 11.73 citations per year, and 1.36 normalized total citations.

XSL•FO
RenderX

Table . Most globally cited studies on machine learning in HIV testing.

Paper details (first author, year, journal)	Digital object identifier (DOI)	TC[a]	TC per year	Normalized TC
Schnall R, 2014, *J Med Internet Res J*	10.2196/jmir.3393	129	11.73	1.36
Zou HC, 2017, *Arch Sex Behav*	10.1007/s10508-016-0709-3	122	15.25	3.40
Odeny TA, 2014, *AIDS*	10.1097/QAD.0000000000000409	116	10.55	1.22
Chamie G, 2016, *Lancet HIV*	10.1016/S2352-3018(15)00251-9	115	12.78	2.37
Rendina HJ, 2014, *AIDS Behav*	10.1007/s10461-013-0573-2	114	10.36	1.20
Jaiswal J, 2019, *Behav Med*	10.1080/08964289.2019.1630357	106	17.67	4.18
Tang WM, 2018, *PLOS Med*	10.1371/journal.pmed.1002645	101	14.43	3.42
Bien CH, 2015, *AIDS Behav*	10.1007/s10461-014-0994-6	97	9.70	2.58
Sullivan PS, 2017, *JMIR mHealth uHealth*	10.2196/mhealth.7199	96	12.00	2.68
Marcus JL, 2019, *Lancet HIV*	10.1016/S2352-3018(19)30137-7	87	14.50	3.43

[a]TC: total citations.

This bibliometric analysis further examines the relevance of institutions and countries worldwide in terms of scientific productivity using publication metrics. Most of the articles included in this study were published by the University of California system (n=67), followed by the University of North Carolina (n=57), the University of North Carolina Chapel Hill (n=52), and the University of Washington (n=48; Table 6). Johns Hopkins University published the fewest studies among the top 10 institutions. As for countries' productivity, authors' correspondence information was used, with a total of 30 countries in the sample (Multimedia Appendix 1 ["Most Relevant Countries"] and Table 7). Of the 266 articles within the sample, the United States produced 59% (n=157) of the studies, ranking first (Table 7). There was a notable synchronicity between the United States and other countries regarding the frequency of publications of studies using machine learning in HIV testing from 2000 to 2024. After the United States, the next most publications were from China (n=27, 10.2%), South Africa (n=16, 6%), the United Kingdom (n=10, 3.8%), and Australia (n=7, 2.6%). The top 5 countries' scientific production in this field was relatively low before 2013 and steadily increased through 2024, as demonstrated in Figure 3. The United States outperformed all other countries, showing positive trends in the publications of studies on machine learning and health technologies in HIV testing from 2000 to 2024.

Table . Top 10 most relevant institutions by number of publications.

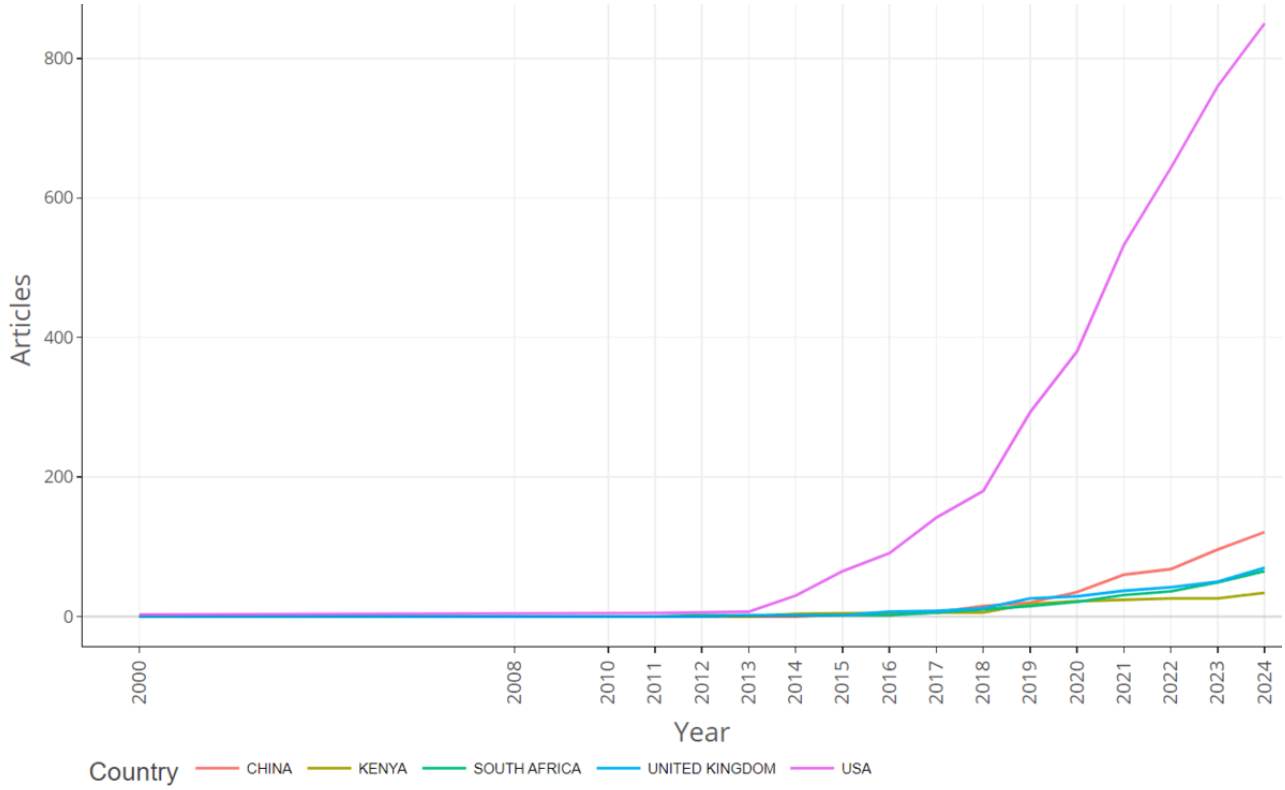
Institution	Articles, n (%)
University of California system	67
University of North Carolina	57
University of North Carolina Chapel Hill	52
University of Washington	48
University of Washington Seattle	48
Emory University	40
Yale University	35
University of Michigan	34
University of Michigan system	34
Johns Hopkins University	33

XSL•FO
RenderX

Table . Top 10 countries by scientific production.

Rank	Country	Articles, n (%)
1	United States of America	157 (59)
2	China	27 (10.2)
3	South Africa	16 (6)
4	United Kingdom	10 (3.8)
5	Australia	7 (2.6)
6	Canada	5 (1.9)
7	Switzerland	5 (1.9)
8	Kenya	4 (1.5)
9	India	3 (1.1)
10	The Netherlands	3 (1.1)

Figure 3. Top 5 countries by annual scientific production. USA: United States of America.



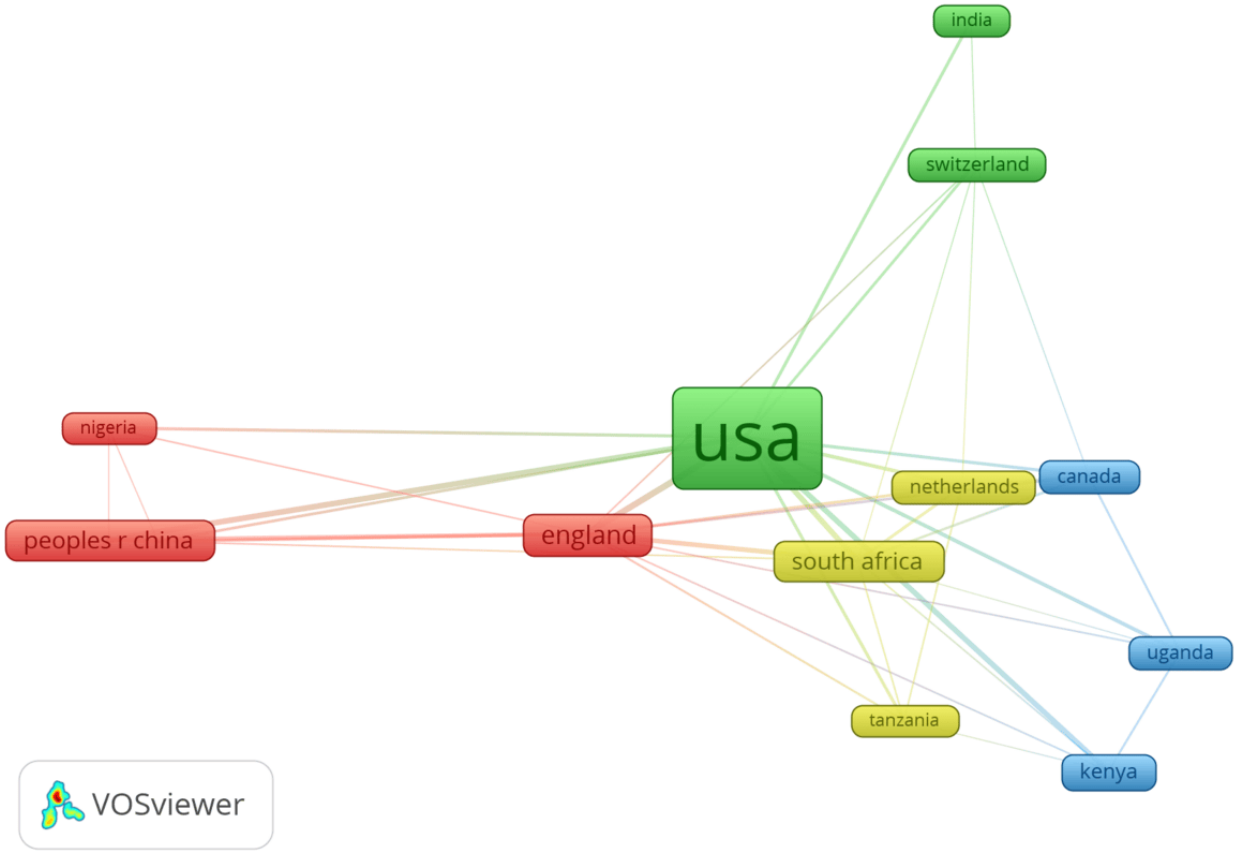
Science Mapping

International Collaboration Networks

Scientific collaborations between countries, institutions, and authors are shown at different levels. Figure 4 presents the most productive countries and their collaboration patterns when publishing studies focusing on machine learning in HIV testing. The country collaboration network shown in Figure 4 consisted of 13 productive countries, each with at least 5 documents. The network was divided into 6 clusters, wherein Canada, Rwanda, Switzerland, and Uganda are in cluster 1 (red). Cluster 2 (green) combines European and African countries such as Germany, the Netherlands, South Africa, and Tanzania. Australia, England, Nigeria, and China are in cluster 3 (blue), and Kenya, Peru, Thailand, and Zimbabwe are in cluster 4 (yellow). Cluster 5 (violet) comprises only European countries, including North Ireland, Norway, and Sweden. Cluster 6 (light blue) has 3 countries from different continents: Brazil, Malaysia, and the United States. The United States emerged as the most productive country, with a link strength of 118, and partnered with 100% of the countries, followed by England, which partnered with 77% of the 13 countries in the network. Australia and India have the lowest amount of country collaboration in studies applying machine learning and emerging health technologies in HIV testing (Figure 4).

XSL•FO
RenderX

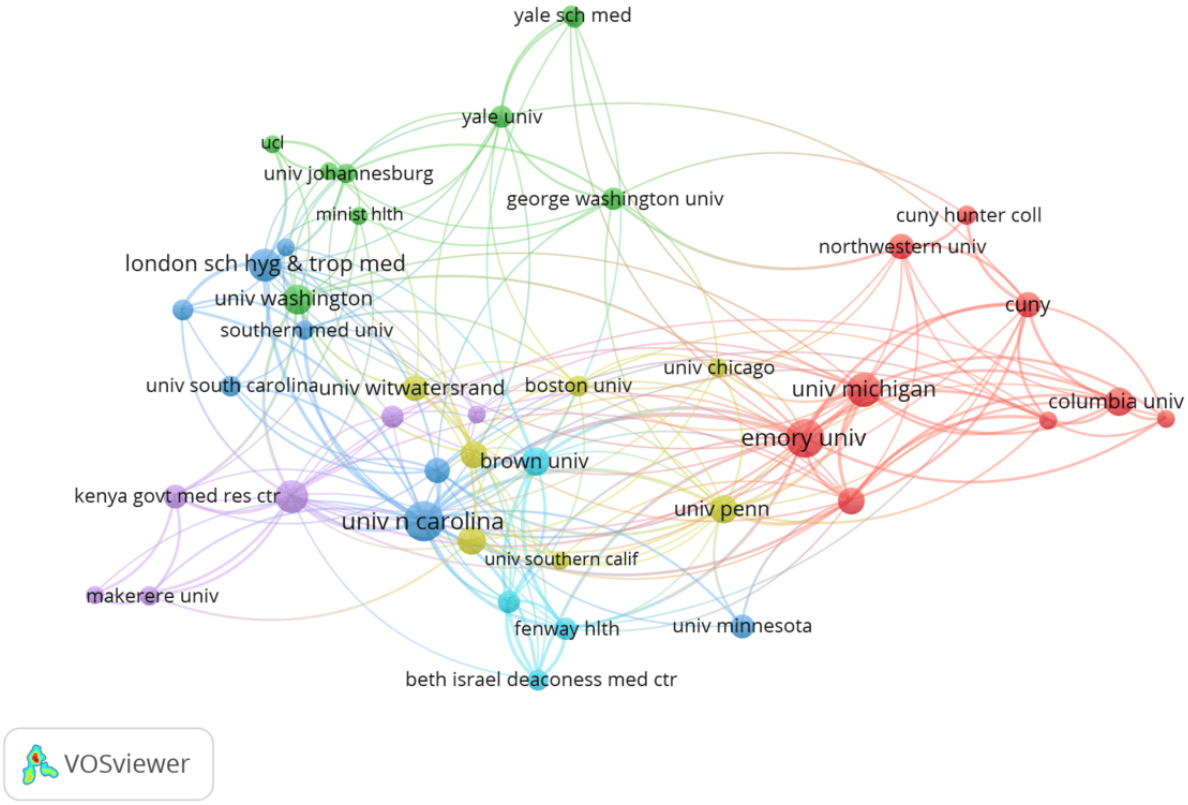
Figure 4. Country collaboration network. Peoples R China: People's Republic of China; USA: United States of America.



Regarding the institutional collaboration network, each institution was set to include at least 5 documents, resulting in 43 productive universities (Figure 5). The most productive universities are in the United States. In terms of link strengths, the University of North Carolina ranked the highest (n=77), with 922 citations, followed by Emory University (n=62), the University of Michigan (n=45), San Diego State University (n=41), and the University of Pennsylvania (n=38). The University of Witwatersrand in South Africa had the highest link strength (n=7) among the African universities, with 158 citations. Among the top 16 universities, the University of North Carolina, Emory University, the University of Pennsylvania, and the University of Michigan have each collaborated with >60% of the institutions within the network. The only African university with good performance is the University of Witwatersrand, exhibiting 19% coauthorship within the network of 43 best universities writing papers on machine learning and emerging health technologies in HIV testing.

XSL•FO
RenderX

Figure 5. Institutional collaboration network.

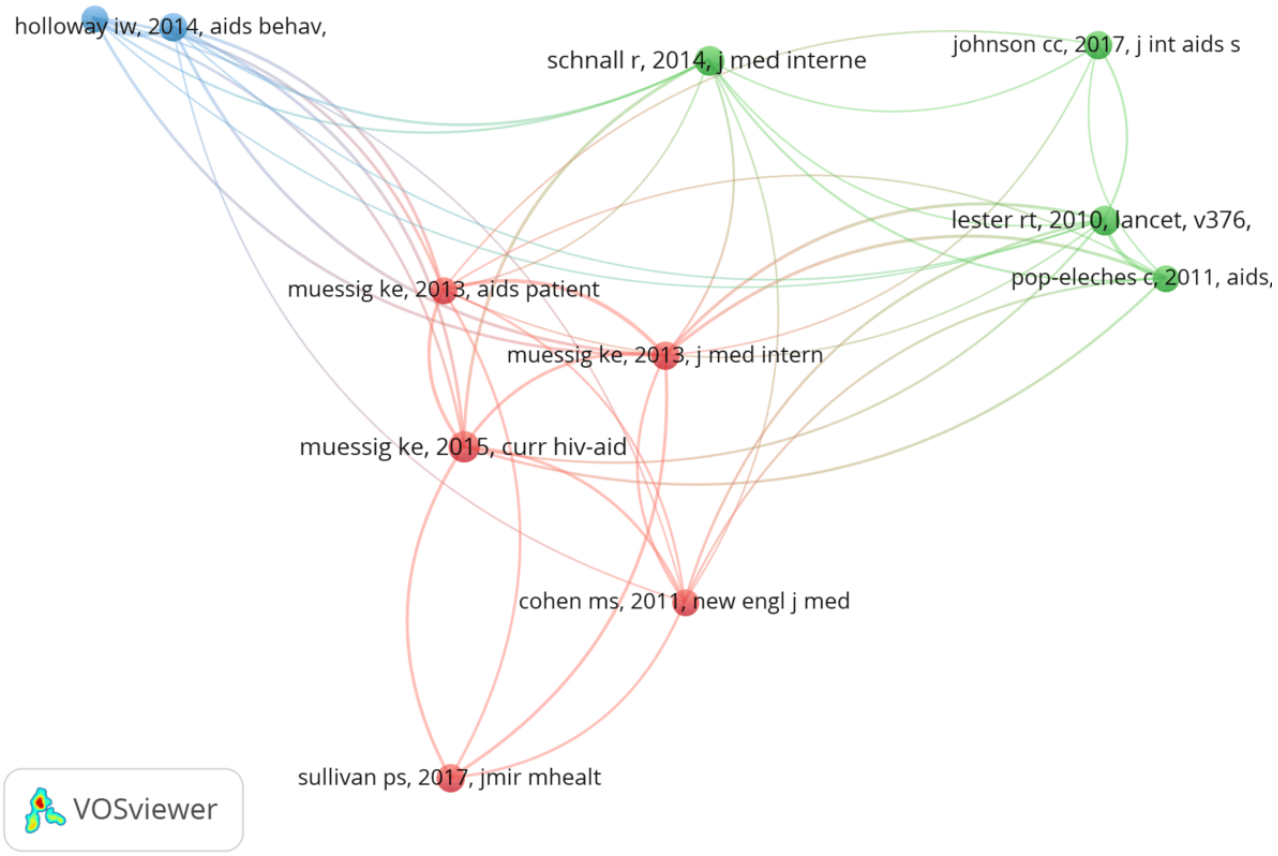


Co-Citation Analysis

A network of co-citation analysis in Figure 6 shows the relationship among cited publications within the research area. The network was set to include only articles with a minimum of 15 citations, in which 11 documents met the threshold. Each article within the network is represented by a node, and the links between the nodes indicate co-cited studies, wherein the studies' strengths and their co-citations are proportional to the nodes' sizes and links' thicknesses, respectively. The most cited document within the network was published by Muessig et al in 2015 [28] (25 citations), titled "A Systematic Review of Recent Smartphone, Internet, and Web 2.0 Interventions to Address the HIV Continuum of Care." The study synthesized evidence from observational and experimental studies focused on understanding the dynamics of technology toward the HIV care cascade, including HIV testing. Three thematic clusters were produced within the network map of co-cited references (Figure 6). Cluster 1 (red) relates to studies on applying machine learning (technology) in HIV prevention in the context of testing. Cluster 2 (green) is associated with studies using machine learning (technology) in HIVST, and the last (blue) cluster within the network focused on the application of mHealth to improve HIV testing among men who have sex with men (MSM). The most globally cited documents are summarized in Table 5 with their digital object identifiers (DOI).

XSL•FO
RenderX

Figure 6. Network of co-cited references.



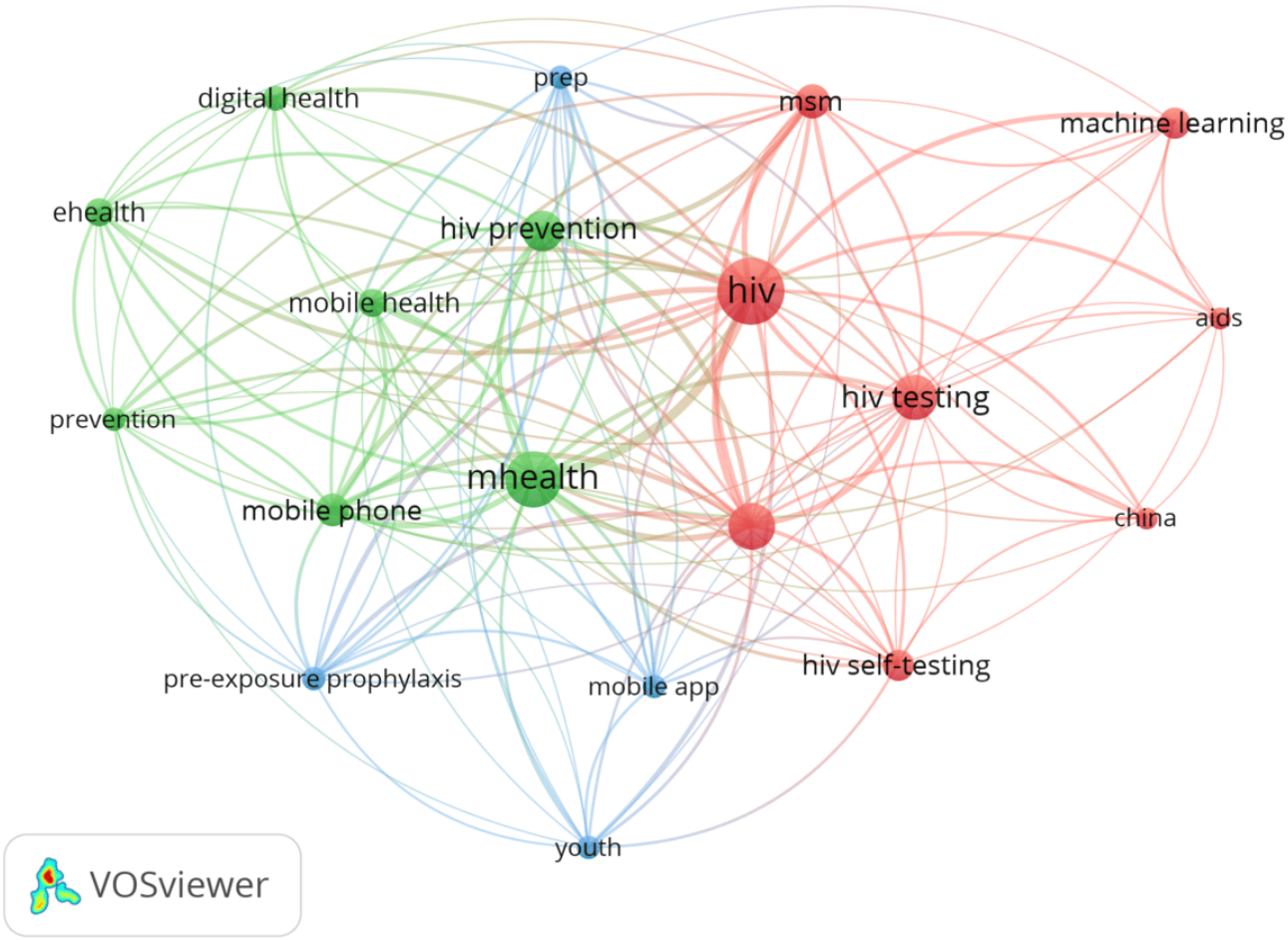
Occurrence of Authors' Keywords

The analysis presents the most frequent keywords used by authors applying machine learning and emerging health technologies in HIV testing. The word cloud in Multimedia Appendix 2 shows the authors' 50 most frequent keywords in this research domain. The word's size in the cloud is proportional to the frequency at which it appears in the bibliographic sample. From the 50 frequently used keywords, HIV ranked the highest, with an occurrence of 93, followed by mHealth (n=65), MSM (n=46), HIV testing (n=41), HIV prevention (n=34), mobile phone (n=22), HIVST (n=21), and machine learning (n=20). These words established the thematic focus of the content analysis.

Figure 7 shows a network of co-occurrence of the authors' most frequent keywords, providing an understanding of their interrelatedness. Using a threshold of 10 for the number of keyword occurrences, a network of the 19 most relevant keywords highlights that HIV has the highest link strength (n=177), with 100% co-occurrence of the 19 keywords. The following keywords also exhibit high link strengths in co-occurrences: mHealth, HIV prevention, MSM, HIV testing, and machine learning. A total of 3 clusters consisting of words within closely related topics were derived from the 19 author keywords within the co-word analysis (Figure 7). Cluster 1 (red) is related to the use of machine learning in HIV testing among MSM, with HIV, AIDS, HIV testing, HIVST, and MSM being the top keywords. Cluster 2 (green) is associated with mHealth and HIV prevention–related topics, and the top keywords are mHealth, digital health, mobile health, and HIV prevention. Likewise, the third cluster (blue), which encompasses keywords such as mobile app, youth, and pre-exposure prophylaxis, also relates to digital health and HIV prevention interventions among youths.

XSL•FO
RenderX

Figure 7. Co-occurrence network.



Content Analysis

This bibliometric analysis explores studies that apply different machine learning techniques or technologies, including eHealth and mHealth, to improve HIV testing and provide insights into the research landscape. We reviewed the contents of 12 articles, with the most prominent authors' keywords presented in Figure 7. The content analysis qualitatively summarized the key findings of the reviewed studies, focusing on the main concepts and topics constructed by the word cloudand co-occurrences (Figure 7) of authors' keywords.

The Application of Technology in HIV Prevention Among Priority Populations

Priority populations consist of young people and the different categories of KPs who are highly susceptible to HIV [29]. Based on the themes generated in the bibliometric analysis, we focus mainly on young people and MSM as key drivers of HIV due to societal stigma and discrimination, lack of HIV knowledge, and engaging in risky behaviors [29]. On top of that, their HIV testing uptake remains low, increasing the spread of the virus and the number of people living with HIV with unknown status [29]. Priority populations' vulnerability to HIV makes them eligible for special interventions, and a variety of emerging technologies are being used to provide efficient and effective HIV prevention methods. We explored different studies addressing the gap within this thematic area. A machine learning technique predicted the internet health information–seeking behaviors of youths at risk of acquiring HIV by using their self-reported internet use data [30]. The study found that about 50% of the youths sought general health care information and sexual and reproductive health information, while 25% sought social services information [30]. In the same study, cisgender and transgender youths were less likely to seek health care information [30]. In another study conducted to predict HIV status among MSM in Zimbabwe, the authors applied different machine learning algorithms [31]. The findings showed that they could be used to enhance the early detection of HIV in order to establish targeted HIV testing for such hard-to-reach populations [31]. The two studies highlight the impact of societal stigma and that a lack of health care information impedes the success of HIV prevention and testing among young people and KPs and suggest the adoption of personalized HIV counseling and testing services [30,31]. Several machine learning and digital technologies, including those encompassed by eHealth and mHealth, have provided efficient HIV testing services by overcoming barriers related to privacy concerns among priority populations [32-37]. From identifying the most eligible candidates for HIV testing to enabling HIVST devices, these studies demonstrate how technology advancements have improved HIV prevention in the context of testing over the past 2 decades [32-37].

XSL•FO
RenderX

Enhancing HIV Self-Testing Using Machine Learning and mHealth Solutions

Self-testing enables people to test for HIV and get their results at home or privately, offering an opportunity to overcome privacy concerns associated with HIV testing, especially among KPs [38]. This is a promising approach to scale up HIV testing but also raises ethical concerns regarding HIV counseling and testing. Advancements in machine learning and mHealth technologies have been used to address this challenge. A feasibility and acceptability assessment of HIVST using conversational agents in South Africa showed substantial progress in this approach [39]. The study developed chatbots for smartphones that were able to give written informed consent to the study participants, allowing them to self-counsel and test for HIV [39]. Fascinatingly, 79% of the users stated that their experiences with the chatbot were preferable to a human counselor [39]. A pragmatic trial involving refugee adolescents and youth in Uganda examined the feasibility of HIVST with and without mHealth [40]. The study participants, who were cisgender and transgender men and women, exhibited increased self-reported HIV testing uptake with the mHealth HIVST approach [40]. Another study [41] also emphasized the critical role mobile phones and digital technology can play in improving HIVST and linkage to care. Despite the opportunity to combat HIV testing barriers among KPs, a review of technology, policy, and evidence regarding HIVST in SSA reported that there is limited delivery within the region [42]. Most of these studies were piloted in Southern and Eastern Africa, necessitating future feasibility studies of HIVST in other African subregions.

Discussion

Principal Findings

This bibliometric analysis provides comprehensive insights into the evolution of machine learning and related technologies in HIV testing interventions from 2000 to 2024. By examining the trends and patterns of research outputs in this field over the past 2 decades, the study shows that publications were relatively low before 2010. The number of publications applying machine learning and emerging health technologies in HIV testing increased rapidly starting in 2013 and was highest in 2021, indicating research progress in this field during this period. The advancement of technology, among other factors, has contributed to the recent rapid growth of publications in this field. Recent breakthroughs in AI have led to increased acceptance in health care [43]. In the past decade, robotics, speech recognition, and machine learning have created new potential for disease prediction, diagnosis, and treatment [44,45]. Machine learning research in health care is expected to grow exponentially, leading to more publications in the future [46]. Our findings are consistent with a bibliometric analysis demonstrating the current growth of AI in health care research [46], and the authors project that the publication rate will double in the next 5 years. Moreover, the surge of publications in different public health domains, including HIV, especially in 2021, was driven by the COVID-19 pandemic, and our topic is no exception [47].

This study showed that the geographical distribution of the scientific production (articles) and collaboration were unequal. The United States ranked the highest, followed by China, South Africa, the United Kingdom, and Australia. These countries are home to top global universities and are well-known for their remarkable advancements in technology-driven medical research. Our analysis further emphasizes that academic collaborations on machine learning and health technologies in HIV testing are more common among universities in high-income countries. The University of North Carolina, Emory University, the University of Michigan, San Diego State University, and the University of Pennsylvania, all within the United States, exhibited the highest collaboration. The only African university that performed well was the University of Witwatersrand in South Africa. Unfortunately, low-income countries, which are disproportionately affected by the HIV epidemic, are missing out on the opportunity to leverage machine learning to enhance HIV testing. A similar study by Amusa et al [23] titled “Big Data and Infectious Disease in Epidemiology: Bibliometric Analysis and Research Agenda” revealed that institutions in advanced countries such as the United States and China contribute the most publications to this field. Previous studies also demonstrated increased academic collaboration among institutions in high-income countries [21,23,48], highlighting the need for greater equity in research partnerships.

The lack of collaboration within and across different settings may jeopardize knowledge translation into evidence-based programs and policies. This increases the need for support from the most productive countries (United States, European countries) to enhance the application of evidence on machine learning and emerging technologies in HIV testing in different contexts (eg, in SSA). Hence, productive universities and/or initiatives within SSA countries should foster collaborative research across different regions and continents in the future.

Subjects related to machine learning, technology, and HIV testing are usually published in public health–related journals, as shown in the analysis. The *Journal of AIDS and Behavior* recorded the highest h-index (10), with 494 citations. *The Journal of Medical Internet Research*, *JMIR mHealth and uHealth*, *JMIR Research Protocols*, *mHealth*, *AIDS Care-Psychological and Socio-Medical Aspects of AI*, *BMC Public Health*, and *PLOS One* were also very influential in publishing studies on machine learning and health technologies in HIV testing. Similar findings were reported in bibliometric analyses focused on digital technologies and HIV research [21,22,48,49].

In analyzing the authors’ keywords, we identified themes and topics based on the frequency and co-occurrence of specific words. Keywords that appeared together within the same cluster indicated related topics. Frequent keywords such as HIV, HIV prevention, MSM, HIVST, HIV testing, mHealth, mobile app, youth, digital health, eHealth, and machine learning appeared in almost all the clusters in the co-word analysis, resulting in 2 thematic topics. These themes reflect the current focus of this field and suggest future research on machine learning and technology in connection with HIV testing. The content analysis, which focused mainly on the application of machine learning

XSL•FO
RenderX

and digital technologies in HIV prevention, including HIVST among KPs, gave comprehensive insights into the extent to which such technologies have enhanced HIV testing and identified limitations as well as suggestions for future studies. The analysis revealed that machine learning and digital technologies can significantly enhance HIV testing among priority groups, including young people and KPs like transgender people, cisgender people, and MSM. Key populations are becoming the center of focus when designing and implementing HIV prevention interventions [29-31]. Such groups are hard to reach due to societal stigma and legal issues, especially in SSA countries [29-31].

A recent systematic review highlighted that machine learning has significant potential to address privacy and accessibility gaps in HIV testing, particularly among KPs [50]. Although machine learning has proven to be highly accurate and effective in HIV testing, the study indicated that most interventions have focused on the Americas, as well as East and Southern Africa [50]. Consistently, our findings showed that machine learning has been used to predict health-seeking behaviors of high-risk populations [30], predict their HIV risk [31], and develop various options aimed at overcoming barriers to KPs accessing and using HIV testing. Among the developments, HIVST was identified as a viable option for conducting self-counseling and testing for HIV among KPs [38-42]. However, feasibility studies for machine learning and emerging technologies in HIV testing approaches have been piloted in only a few Southern and Eastern African countries. Furthermore, lack of expertise and inadequate infrastructure limit the application of advanced technologies and AI in HIV testing across many African countries [50]. This gap could leave the continent behind in meeting the UNAIDS 95-95-95 target despite bearing the brunt of the global HIV burden. Thus, these innovative techniques are essential for strengthening and should be implemented across SSA beyond piloting purposes.

Limitations

This bibliometric analysis encountered several limitations. First, the WoS database was used to gather articles, which might have excluded relevant studies from other databases. The WoS is a leading database for studies on machine learning and technology in health care and has been recognized as a legitimate platform for bibliometric analysis [18-23]. Second, we only selected studies published in English, potentially missing relevant studies published in other languages. Third, our sample size was relatively small and may not fully represent research outputs in machine learning and technology applications in HIV testing. However, our findings are relevant as they are consistent with publication trends in similar bibliometric analyses. Moreover, the Biblioshiny software could not determine authors' valid affiliations; a single author could have multiple affiliations in some instances. We chose the "Application Name Disambiguation" option to ensure the same affiliations were counted as one.

Future Research and Conclusions

This study showed that universities in high-income countries like the United States, the United Kingdom, and South Africa are very productive in applying machine learning technology to address health care issues, including HIV testing. Journals publishing articles on public health and infectious diseases were very influential in this analysis. The most frequent keywords in the sample were related to the application of digital technologies in HIV prevention, particularly HIV testing among KPs. Furthermore, the study found that incorporating machine learning in HIVST seems promising for the enhancement of HIV testing among KPs, especially MSM. These insights demonstrate improved academic engagement and productivity over the 2 decades, identifying opportunities and research gaps for future scholars to embark on.

This bibliometric analysis highlights a research gap between high-income and underresourced nations regarding applying machine learning and related technologies in HIV testing interventions. More than half of the studies analyzed were from the United States, and a handful of them came from other advanced countries. Many low-income countries, especially in SSA, are lagging in using advanced technologies to tackle health care challenges. There is a critical need to develop an enabling infrastructure that can facilitate the use of machine learning in HIV testing interventions. Furthermore, academic institutions should incorporate machine learning courses to equip future researchers with the essential skills needed in this area. Bridging the intercontinental gap is crucial for overcoming the barriers to HIV prevention interventions, ultimately contributing to the success of UNAIDS' goals to eradicate HIV/AIDS as a global health issue.

This bibliometric analysis clearly demonstrates a significant shift from conventional HIV testing methods to innovative approaches like HIVST. The WHO strongly endorses HIVST as an effective way to deliver personalized and context-specific HIV services that tackle the structural issues surrounding stigma and discrimination. Notably, both youth and KPs have expressed a clear preference for HIVST and mobile technology, which are crucial for improving linkage to care. Although HIVST is user-friendly and allows individuals to test conveniently at home, it presents a challenge in accurately monitoring national HIV testing rates. To address this issue, it is imperative that HIV policies implement robust tracking systems to ensure that individuals using home testing are included in national HIV testing data. This proactive approach is essential for effectively assessing and improving HIV care and services.

Machine learning models continually evolve to address stigma and discrimination, as well as privacy concerns related to HIV testing, proving to be efficient and effective despite implementation challenges. Future studies should assess the feasibility of various machine learning techniques to enhance HIV testing in low-income countries. Such insights will further identify solutions to the current gaps in research in low-income countries, particularly those in SSA. The study also highlights the need to improve collaboration among African universities to enhance research outputs using advanced methodologies. In today's world, utilizing AI, machine learning, and advanced technologies is essential for addressing critical challenges across all sectors, especially health care. Based on the gaps and opportunities highlighted in this study, the findings would inform program implementers to design more effective strategies for improving the use of machine learning and technology in

XSL•FO
RenderX

HIV testing. By increasing the use of machine learning and emerging technologies in HIV testing interventions, we can significantly boost the number of individuals who get tested for HIV. This will contribute to the success of the UNAIDS 2030 goal and result in improved health outcomes for our communities.

Acknowledgments

This study is part of the Boloka Project, which is funded by the South African Medical Research Council (SAMRC) under its Division of Research Capacity Development through the Mid-Career Scientist Programme, using funds allocated by the South African National Treasury. The authors, namely RNPM, EP, and MJ, are supported by the SAMRC. This study was conducted within the SAMRC/University of Johannesburg (UJ) Pan African Centre for Epidemics Research Extramural Unit framework. The bibliometric analysis also forms part of a doctoral study by MJ, whose study is funded by the UJ Global Excellence Stature 4.0 Scholarship. The authors are solely responsible for the views expressed in this work, and they do not necessarily reflect the official perspectives of the SAMRC or UJ.

Authors' Contributions

M.J., R.N.P.M., and E.P. conceived the study and reviewed the literature. M.J. and E.P. screened articles. M.J. wrote the first draft of the paper. M.J. and A.L.B. analyzed and visualized the data. R.N.P.M., E.P., Y.A.S., A.L.B., and H.T. reviewed the work. R.N.P.M., E.P., and Y.A.S. supervised the study. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Most relevant countries.
[XLSX File, 4 KB - i-jmr_v14i1e64829_app1.xlsx]

Multimedia Appendix 2
Word cloud of authors' keywords.
[PNG File, 76 KB - i-jmr_v14i1e64829_app2.png]

References

1. The global health observatory: summary of the global HIV endemic, 2022. World Health Organization. URL: https://www.who.int/data/gho/data/themes/hiv-aids [accessed 2025-05-12]
2. Global HIV & AIDS statistics. UNAIDS. URL: https://www.unaids.org/en/resources/fact-sheet [accessed 2025-05-12]
3. 2025 target: putting people living with HIV and communities at risk at the centre. UNAIDS. URL: https://aidstargets2025.unaids.org [accessed 2025-05-12]
4. Global HIV programme: HIV testing services. World Health Organization. URL: https://www.who.int/teams/global-hiv-hepatitis-and-stis-programmes/hiv/testing-diagnostics/hiv-testing-services [accessed 2025-05-12]
5. WHO recommends self-HIV testing – evidence update and considerations for success. World Health Organization. URL: https://www.who.int/publications/i/item/WHO-CDS-HIV-19.36 [accessed 2025-05-12]
6. Romero RA, Klausner JD, Marsch LA, Young SD. Technology-delivered intervention strategies to bolster HIV testing. Curr HIV/AIDS Rep 2021 Aug;18(4):391-405. [doi: 10.1007/s11904-021-00565-y] [Medline: 34109549]
7. Mutai CK, McSharry PE, Ngaruye I, Musabanganji E. Use of machine learning techniques to identify HIV predictors for screening in sub-Saharan Africa. BMC Med Res Methodol 2021 Jul 31;21(1):159. [doi: 10.1186/s12874-021-01346-2] [Medline: 34332540]
8. Dubey A. Biophotonics and machine learning model for the diagnosis and treatment of HIV. Biosci Biotech Res Comm 2018 Jan 25;11(1):5-10 [FREE Full text] [doi: 10.21786/bbrc/11.1/2]
9. Turk G, Ghiglione Y, Hormanstorfer M, et al. Biomarkers of progression after HIV acute/early infection: nothing compares to CD4. Viruses 2018 Jan 13;10(1):1-24. [doi: 10.3390/v10010034] [Medline: 29342870]
10. Chikusi H, Leo J, Kaijage S. Machine learning model for prediction and visualization of HIV index testing in Northern Tanzania. IJACSA 2022;13(2):391-399. [doi: 10.14569/IJACSA.2022.0130246]
11. Xu X, Ge Z, Chow EPF, et al. A machine-learning-based risk-prediction tool for HIV and sexually transmitted infections acquisition over the next 12 months. J Clin Med 2022 Mar 25;11(7):2016-2021. [doi: 10.3390/jcm11071818] [Medline: 35407428]
12. Pan Y, Liu H, Metsch LR, Feaster DJ. Factors associated with HIV testing among participants from substance use disorder treatment programs in the US: a machine learning approach. AIDS Behav 2017 Feb;21(2):534-546. [doi: 10.1007/s10461-016-1628-y] [Medline: 27933461]

XSL•FO
RenderX

13. Lu Y, Ni Y, Wang Q, et al. Effectiveness of sexual health influencers identified by an ensemble machine learning model in promoting secondary distribution of HIV self-testing among men who have sex with men in China: study protocol for a quasi-experimental trial. BMC Public Health 2021 Sep 28;21(1):1772. [doi: 10.1186/s12889-021-11817-2] [Medline: 34583667]
14. van Heerden A, Ntinga X, Vilakazi K. The potential of conversational agents to provide a rapid HIV counseling and testing services. Presented at: 2017 International Conference on the Frontiers and Advances in Data Science (FADS); Oct 23-25, 2017; Xi'an, China. [doi: 10.1109/FADS.2017.8253198]
15. Turbé V, Herbst C, Mngomezulu T, et al. Deep learning of HIV field-based rapid tests. Nat Med 2021 Jul;27(7):1165-1170. [doi: 10.1038/s41591-021-01384-9] [Medline: 34140702]
16. Gupta K, Ruan Y, Ibrahim A, et al. Transforming rapid diagnostic tests into trusted diagnostic tools in LMIC using AI. Presented at: 2023 IEEE Conference on Artificial Intelligence (CAI); Jun 5-6, 2023; Santa Clara, CA. [doi: 10.1109/CAI54212.2023.00136]
17. Roche SD, Ekwunife OI, Mendonca R, et al. Measuring the performance of computer vision artificial intelligence to interpret images of HIV self-testing results. Front Public Health 2024;12:1334881. [doi: 10.3389/fpubh.2024.1334881] [Medline: 38384878]
18. Web of Science. 2024. URL: https://www.webofscience.com/wos/woscc/basic-search
19. Klarin A. How to conduct a bibliometric content analysis: guidelines and contributions of content co‐occurrence or co‐word literature reviews. Int J Consumer Studies 2024 Mar;48(2):1-20. [doi: 10.1111/ijcs.13031]
20. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res 2021 Sep;133(March):285-296. [doi: 10.1016/j.jbusres.2021.04.070]
21. Mhando F, Nyankomo M, Hall C, et al. Digital intervention services to promote HIV self-testing and linkage to care services: a bibliometric and content analysis-global trends and future directions. Public Health Rev 2024;45(February):1606354. [doi: 10.3389/phrs.2024.1606354] [Medline: 38434540]
22. Wang N, Zhang R, Ye Z, et al. Studies on HIV/AIDS among students: bibliometric analysis. Interact J Med Res 2023 Aug 4;12:e46042. [doi: 10.2196/46042] [Medline: 37540553]
23. Amusa LB, Twinomurinzi H, Phalane E, Phaswana-Mafuya RN. Big data and infectious disease epidemiology: bibliometric analysis and research agenda. Interact J Med Res 2023 Mar 31;12:e42292. [doi: 10.2196/42292] [Medline: 36913554]
24. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. J Informetr 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]
25. Lim WM, Kumar S. Guidelines for interpreting the results of bibliometric analysis: a sensemaking approach. Glob Bus Org Exc 2024 Jan;43(2):17-26. [doi: 10.1002/joe.22229]
26. Hirsch JE. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. Scientometrics 2010 Dec;85(3):741-754. [doi: 10.1007/s11192-010-0193-9]
27. Schnall R, Travers J, Rojas M, Carballo-Diéguez A. eHealth interventions for HIV prevention in high-risk men who have sex with men: a systematic review. J Med Internet Res 2014 May 26;16(5):e134. [doi: 10.2196/jmir.3393] [Medline: 24862459]
28. Muessig KE, Nekkanti M, Bauermeister J, Bull S, Hightow-Weidman LB. A systematic review of recent smartphone, Internet and Web 2.0 interventions to address the HIV continuum of care. Curr HIV/AIDS Rep 2015 Mar;12(1):173-190. [doi: 10.1007/s11904-014-0239-3] [Medline: 25626718]
29. Consolidated guidelines on HIV prevention, diagnosis, treatment and care for key populations. World Health Organization. URL: http://apps.who.int/iris/bitstream/10665/128048/1/9789241507431_eng.pdf?ua=1 [accessed 2025-05-12]
30. Comulada WS, Goldbeck C, Almirol E, et al. Using machine learning to predict young people's internet health and social service information seeking. Prev Sci 2021 Nov;22(8):1173-1184. [doi: 10.1007/s11121-021-01255-2] [Medline: 33974226]
31. Chingombe I, Dzinamarira T, Cuadros D, et al. Predicting HIV status among men who have sex with men in Bulawayo & Harare, Zimbabwe using bio-behavioural data, recurrent neural networks, and machine learning techniques. Trop Med Infect Dis 2022 Sep 5;7(9):231. [doi: 10.3390/tropicalmed7090231] [Medline: 36136641]
32. Peng ML, Wickersham JA, Altice FL, et al. Formative evaluation of the acceptance of HIV prevention artificial intelligence chatbots by men who have sex with men in Malaysia: focus group study. JMIR Form Res 2022 Oct 6;6(10):e42055. [doi: 10.2196/42055] [Medline: 36201390]
33. Janssen R, Engel N, Pant Pai N, et al. 'You're only there on the phone'? A qualitative exploration of community, affect and agential capacity in HIV self-testing using a smartphone app. Sociol Health Illn 2021 Mar;43(3):591-606. [doi: 10.1111/1467-9566.13242] [Medline: 33634889]
34. Menza TW, Hughes JP, Celum CL, Golden MR. Prediction of HIV acquisition among men who have sex with men. Sex Transm Dis 2009 Sep;36(9):547-555. [doi: 10.1097/OLQ.0b013e3181a9cc41] [Medline: 19707108]
35. Zhao Y, Zhu X, Pérez AE, et al. MHealth approach to promote oral HIV self-testing among men who have sex with men in China: a qualitative description. BMC Public Health 2018 Sep 27;18(1):1146. [doi: 10.1186/s12889-018-6046-9] [Medline: 30261856]

XSL•FO
RenderX

36. Prata Menezes N, Mehta SH, Wesolowski A, et al. Network centrality and HIV prevention service use among people who inject drugs: findings from a sociometric network cohort in New Delhi, India. Addiction 2024 Mar;119(3):570-581. [doi: 10.1111/add.16379] [Medline: 37967827]
37. Wang B, Liu F, Deveaux L, et al. Adolescent HIV-related behavioural prediction using machine learning: a foundation for precision HIV prevention. AIDS 2021 May 1;35(Suppl 1):S75-S84. [doi: 10.1097/QAD.0000000000002867] [Medline: 33867490]
38. HOME-BASED HIV SELF-TESTING WITH ONLINE INSTRUCTION AND COUNSELING (HIVST-OIC). Centers for Disease Control and Prevention. URL: https://stacks.cdc.gov/view/cdc/149596 [accessed 2025-05-12]
39. Ntinga X, Musiello F, Keter AK, Barnabas R, van Heerden A. The feasibility and acceptability of an mHealth conversational agent designed to support HIV self-testing in South Africa: cross-sectional study. J Med Internet Res 2022 Dec 12;24(12):e39816. [doi: 10.2196/39816] [Medline: 36508248]
40. Logie CH, Okumu M, Berry I, et al. Findings from the Tushirikiane mobile health (mHealth) HIV self-testing pragmatic trial with refugee adolescents and youth living in informal settlements in Kampala, Uganda. J Int AIDS Soc 2023 Oct;26(10):e26185. [doi: 10.1002/jia2.26185] [Medline: 37850816]
41. Muwanguzi PA, Nasuuna EM, Namimbi F, Osingada CP, Ngabirano TD. Venues and methods to improve professional men's access to HIV self-testing and linkage to HIV prevention or treatment: a qualitative study. BMC Health Serv Res 2021 Nov 9;21(1):1217. [doi: 10.1186/s12913-021-07259-6] [Medline: 34753460]
42. Indravudh PP, Choko AT, Corbett EL. Scaling up HIV self-testing in sub-Saharan Africa: a review of technology, policy and evidence. Curr Opin Infect Dis 2018 Feb;31(1):14-24. [doi: 10.1097/QCO.0000000000000426] [Medline: 29232277]
43. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
44. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. Ann Surg 2018 Jul;268(1):70-76. [doi: 10.1097/SLA.0000000000002693] [Medline: 29389679]
45. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med 2019 Jan;25(1):24-29. [doi: 10.1038/s41591-018-0316-z] [Medline: 30617335]
46. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial intelligence in health care: bibliometric analysis. J Med Internet Res 2020 Jul 29;22(7):e18228. [doi: 10.2196/18228] [Medline: 32723713]
47. Chinnery PF, Pearce JJ, Kinsey AM, Jenkinson JM, Wells G, Watt FM. How COVID-19 has changed medical research funding. Interface Focus 2021 Dec 6;11(6):20210025. [doi: 10.1098/rsfs.2021.0025] [Medline: 34956595]
48. Liang C, Qiao S, Olatosi B, Lyu T, Li X. Emergence and evolution of big data science in HIV research: bibliometric analysis of federally sponsored studies 2000-2019. Int J Med Inform 2021 Oct;154(July):104558. [doi: 10.1016/j.ijmedinf.2021.104558] [Medline: 34481301]
49. Rosas SR, Kagan JM, Schouten JT, Slack PA, Trochim WMK. Evaluating research and impact: a bibliometric analysis of research by the NIH/NIAID HIV/AIDS clinical trials networks. PLOS One 2011 Mar 4;6(3):e17428. [doi: 10.1371/journal.pone.0017428] [Medline: 21394198]
50. Jaiteh M, Phalane E, Shiferaw YA, Voet KA, Phaswana-Mafuya RN. Utilization of machine learning algorithms for the strengthening of HIV testing: a systematic review. Algorithms 2024;17(8):362. [doi: 10.3390/a17080362]

Abbreviations

AI: artificial intelligence
DOI: digital object identifier
HIVST: HIV self-testing
KP: key population
MSM: men who have sex with men
PACER: Pan African Centre for Epidemics Research
RF: random forest
SAMRC: South African Medical Research Council
SSA: sub-Saharan Africa
UJ: University of Johannesburg
UNAIDS: The Joint United Nations Programme on HIV/AIDS
WHO: World Health Organization
WoS: Web of Science
XGBoost: Extreme Gradient Boosting

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 28.07.24; peer-reviewed by F Yang, SA Bakar; revised version received 12.12.24; accepted 12.12.24; published 22.05.25.

Please cite as:
Jaiteh M, Phalane E, Shiferaw YA, Amusa LB, Twinomurinzi H, Phaswana-Mafuya RN
Application of Machine Learning and Emerging Health Technologies in the Uptake of HIV Testing: Bibliometric Analysis of Studies Published From 2000 to 2024
Interact J Med Res 2025;14:e64829
URL: https://www.i-jmr.org/2025/1/e64829
doi:10.2196/64829

© Musa Jaiteh, Edith Phalane, Yegnanew A Shiferaw, Lateef Babatunde Amusa, Hossana Twinomurinzi, Refilwe Nancy Phaswana-Mafuya. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.5.2025. 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 Effects of Virtual Reality on Hope and Travel Expectations in Healthy and Hospitalized Children: Quasi-Experimental Design Approach

Pei-Shan Hsieh[1], PhD; Hsiu-Sen Chiang[2], PhD; Fang-Liang Huang[3], PhD; Wen-Hsin Chi[4], PhD

[1]Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, No. 58, Shenzhong Rd, Yanchao Dist, Kaohsiung City, Taiwan

[2]Department of Information Management, National Taichung University of Science, Taichung City, Taiwan

[3]Department of Children's Medical Center, Taichung Veterans General Hospital, Taichung, Taiwan

[4]Institute of Management of Technology, National Yang Ming Chiao Tung University, Hsinchu, Taiwan

Corresponding Author:
Pei-Shan Hsieh, PhD
Department of Intelligent Commerce, National Kaohsiung University of Science and Technology, No. 58, Shenzhong Rd, Yanchao Dist, Kaohsiung City, Taiwan

Abstract

Background: Virtual reality (VR) has become a powerful tool for enhancing the experiences of patients with critical illnesses, particularly hospitalized children with leukemia. Since traveling is nearly impossible for them, St Jude has teamed up with the travel company Expedia to launch "Dream Adventures," a pilot program that offers immersive VR experiences, allowing children to explore new destinations from the comfort of the hospital.

Objective: The aim of this study was to evaluate the pleasurable experience of VR and its impact on healthy and hospitalized children's travel expectations and hope by using electrocardiography (ECG) and questionnaires to enhance research objectivity.

Methods: Participants were children aged 7-18 years, divided into 2 distinct groups: 30 healthy children and 18 hospitalized children with leukemia. Both groups received the same VR intervention and were assessed using a 1-group pretest-posttest design. The questionnaires were designed to assess differences in the children's sense of hope and travel expectations, and their physiological data were collected through ECG.

Results: The results indicated a statistically significant increase in hope levels from pre-VR to post-VR intervention in both healthy children (preintervention: mean 5.83, SD 0.87; postintervention: mean 6.36, SD 0.76; P=.002) and hospitalized children (preintervention: mean 5.51, SD 1.17; postintervention: mean 5.73, SD 1.15; P=.03), as determined by paired samples 2-tailed t tests. Furthermore, an independent samples 2-tailed t test revealed a significant difference in postintervention hope levels between the hospitalized children (mean 5.73, SD 1.15) and healthy children (mean 6.36, SD 0.76; P=.05). Then, we further compared the mean differences in hope scores from preintervention to postintervention. Healthy children demonstrated a greater increase (an increase of 0.53, from 5.83 to 6.36) than the hospitalized children (an increase of 0.22, from 5.51 to 5.73). In terms of physiological responses, ECG indicators such as SD of all normal-to-normal intervals and low-frequency power revealed significant differences in autonomic nervous system activity between the 2 groups. Healthy children exhibited higher sympathetic activation, suggesting greater emotional engagement during the VR experience, whereas hospitalized children demonstrated more attenuated responses. The consistency between physiological data and self-reported measures strengthens the construct validity of the instruments used and enhances the overall reliability of the study findings.

Conclusions: The VR intervention significantly increased hope levels in both healthy children and hospitalized children with leukemia, with a greater improvement observed among healthy participants. Therefore, this study suggests that when designing interventions for hospitalized children, more targeted emotional support strategies should be considered. Future studies are recommended to explore different types of VR content and the medical conditions of hospitalized children.

(Interact J Med Res 2025;14:e65311) doi:10.2196/65311

KEYWORDS

hospitalized children; virtual reality; ECG; hope; travel expectations; electrocardiogram

XSL•FO
RenderX

Introduction

Nowadays, virtual reality (VR) has gradually become a powerful tool for enhancing the patient experience in children's hospitals [1]. Immersive experiences allow individuals to fully engage in realistic environments created through digital technologies or designed settings, fostering deep emotional connections [2]. VR apps initially started with simple graphics for entertainment and research but are now widely used across many professional fields [3]. As VR technology continues to advance, its apps in the medical field are gaining increasing recognition. For instance, VR is being used in various medical domains [4], including medical education [5,6], clinical decision-making [7,8], preoperative planning [9], patient education [10,11], and surgical simulation [12].

As part of VR content creation, 360-degree video is an innovative medium that captures a scene from all directions using specialized cameras, allowing users to explore the environment as if they were physically present [13]. The sense of spatial presence created by 360-degree videos positions them as a form of immersive VR experience [14,15], particularly due to their ability to evoke strong emotions and make users feel surrounded by the content [16,17]. Previous studies in nonhospital settings have shown that VR has significant potential, as it can reduce psychological distress in young people across various situations [18-20] and elicit varying degrees of physical activity intensity in young healthy adults [21].

Additionally, there is a great potential for VR to address the unique medical care needs of hospitalized children [22]. For example, for children battling serious illnesses at St Jude Children's Research Hospital, traveling may seem out of reach. To bring the world to these young patients, St Jude has teamed up with the travel company Expedia to launch "Dream Adventures," a pilot program that offers immersive VR experiences, allowing kids to explore new destinations from the comfort of the hospital [23].

Mitsea et al [24] highlighted that VR-based positive mental imagery training, when combined with reward training, effectively reduced anhedonia while increasing pleasure and self-motivation skills. van Gelderen et al [25] implemented VR-based Eye Movement Desensitization and Reprocessing therapy, where patients walked on a treadmill while engaging with personal trauma-related images and relaxing music. The results indicated that such interventions could effectively reduce feelings of isolation and foster hope. Additionally, Bush et al [26] found that the Virtual Hope Box application enhanced users' hope and emotional regulation skills. Therefore, the aim of our study was to examine the effects of VR on pleasure, hope, and travel expectations in hospitalized children with leukemia and further compare them with healthy children.

Researchers have used videos to elicit specific emotions and collect emotion-related data. More recently, VR has been used to evoke particular emotional responses [27]. VR offers a more objective method for eliciting emotions by minimizing environmental interferences [28]. However, health care providers have perceived certain barriers to using VR, such as concerns about older adults with cognitive impairments [29]. Nevertheless, studies have suggested that VR apps are feasible for use in hospital wards and do not cause significant side effects [30,31]. Self-reported emotion measurements are susceptible to recall bias and cultural influences, whereas electrocardiography (ECG) provides a reliable, real-time evaluation of emotional variations.

Recent studies have demonstrated that ECG is a viable signal for measuring stress [32], and research in the study by Schmidt et al [33] has shown the feasibility of ECG-based stress assessment in a VR environment. Ahmad et al [34] collected ECG data from 9 participants experiencing a VR roller coaster. von Rosenberg et al [35] investigated the impact of mental stress on physiological responses using a wireless ECG. Furthermore, Kalatzis et al [36] examined whether VR scenarios could elicit specific affective states and developed a validated database to predict emotional states (eg, calm vs acute stress) using ECG. Additionally, Scherz et al [2] used VR as a significant stressor, analyzing heart rate and other physiological parameters through ECG. Their results indicated that while using a VR headset influenced heart rate and the root-mean-square of the successive differences in heartbeats, it did not significantly increase driving-related stress.

Therefore, the aim of this study was to evaluate the pleasurable experience of VR and its impact on hospitalized children's travel expectations and sense of hope by using ECG and questionnaires approach. This study incorporated the collection of physiological data through ECG to analyze and corroborate emotional changes in children based on physiological signals.

Methods

Study Design and Setting

This study used a quasi-experimental design to investigate the effects of VR on hope, pleasure, and travel expectations in children. The study included 2 distinct groups of participants: healthy children and hospitalized children with leukemia. Both groups received the same VR intervention and were assessed using a 1-group pretest-posttest design, in which outcome measures were collected immediately before and after the intervention within each group. This design allowed us to examine within-group changes in hope, pleasure, and travel expectations, as well as to compare the differential effects of the VR intervention between the 2 populations.

Participants

The experiment was conducted in 2 stages: a pilot test and a formal test. The purpose of the pilot test was to determine the VR destination. Five domestic tourist destinations (Park Lane by China Metal Products, National Museum of Natural Science, Dadu Blue Highway, Shalu Dream Street, and Dakeng Pink Valentine Bridge) and 6 international tourist destinations (the Pyramids of Egypt, the Eiffel Tower in France, Neuschwanstein Castle in Germany, Venice in Italy, Aphrodisias in Turkey, and the Space Station) were provided as options. Based on the results of the pilot test involving 12 children aged 7 -18 years, the Eiffel Tower was selected as the VR content due to its high ratings for engagement and recognizability. The content consisted of 8 scenes of the tower, captured from various angles and at

XSL•FO
RenderX

different times of day, and participants were able to switch between scenes freely using the VR device.

In the formal test, participants were included 2 distinct groups of participants: healthy children and hospitalized children with leukemia. The hospitalized children with leukemia were recruited from the pediatric ward of Taichung Veterans General Hospital in Taiwan based on inclusion and exclusion criteria recommended by the attending physician. These children often undergo strict regimens of chemotherapy, radiation, or other therapies, which limit their opportunities to travel and experience new places, particularly if they have spent significant time in the hospital. This context likely heightens their desire to engage in virtual travel experiences, making them a key focus of this study. Additionally, the healthy children were recruited from a local cram school, where they were enrolled in regular academic programs and had no known history of serious medical conditions. Recruitment procedures prioritized voluntary participation. The inclusion and exclusion criteria of formal experimental design are summarized in Textbox 1.

Textbox 1. Inclusion and exclusion criteria for participants in the formal study design.

Inclusion criteria

- Hospitalized children: hospitalized children with leukemia aged 7-18 years, in the midstage of treatment with stable vital signs.
- Healthy children: healthy children aged 7-18 years.

Exclusion criteria

- Hospitalized children: children unable to independently use virtual reality equipment or those diagnosed with epilepsy.
- Healthy children: children unable to independently use virtual reality equipment.

Ethical Considerations

All procedures performed in studies involving human participants adhered to the ethical standards of the institutional and national research committee, as well as the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

This study was approved by the Institutional Review Board of Taichung Veterans General Hospital (CF20285A) on December 3, 2020, and was conducted from December 3, 2020, to June 30, 2021. Informed consent was obtained from all participants and their parents or guardians prior to the study, with signed consent forms collected. The consent process involved clearly communicating the study's objectives, procedures, and potential risks to ensure voluntary participation. All data were anonymized to protect participants' privacy.

To show appreciation for participants' involvement, each participant will receive a VR cardboard viewer (valued at approximately US $1.50) upon completing the experiment. This compensation allows them to continue experiencing VR content after the study. It does not affect their right to participate voluntarily or withdraw from the study at any time.

Instruments

This study adopted the pleasure measurement from the study by Bakhuis et al [37]. A higher score indicated that the subject experienced more pleasure from the stimulus as shown in Table 1.

Table . Scale items used to measure pleasure in response to virtual reality stimuli[a].

Items	Scores
1. Unhappy to happy	1 - 7
2. Unsatisfied to satisfied	1 - 7
3. Melancholic to cheerful	1 - 7
4. Despairing to hopeful	1 - 7
5. Bored to surprised	1 - 7

[a]From Schifferstein and Tanudjaja [38].

In addition, we measured children's hope using the scale developed by Snyder et al [39]. The scale was scored on a 7-point Likert scale, with responses ranging from 1 ("definitely disagree") to 7 ("definitely agree"). Travel expectations in this study were based on the anticipated outcomes a person expects to gain while traveling to a destination [40]. The items measuring children's hope and travel expectations are listed in Textbox 2.

XSL•FO
RenderX

Textbox 2. Scale items for measuring children's hope and travel expectations.

Items of measuring children's hope [39]:

1. My life is very good now.
2. I can think of many ways to get the things in life that are most important to me.
3. I am doing just as well as other kids of my age.
4. I have a lot of confidence in myself.
5. When I have a problem, I can come up with lots of ways to solve it.
6. I think the things I have done in the past will help me in the future.
7. Even when others want to quit, I know that I can find ways to solve the problem.
8. I think my life is full of hope.

Items of measuring travel expectations [40]:

1. I can learn new knowledge.
2. I can learn new culture.
3. I can experience different things.
4. I can see the beautiful view.

Procedure

The researcher tested 1 child at a time, with each session lasting approximately 40 minutes. Healthy children were tested individually in a classroom, while hospitalized children were tested in their hospital ward beds, as shown in Figure 1. The study prioritized participant safety during VR use. Potential risks associated with VR, such as motion sickness, visual discomfort, or dizziness, were considered during the design phase. To mitigate these risks, age-appropriate 360-degree VR content with calming visuals was selected. Participants were monitored throughout the session, and breaks were allowed if discomfort occurred.

Figure 1. Hospitalized children were tested in a hospital ward bed.



As the participants were underage and part of a vulnerable group, the attending physician and researcher carefully explained the experiment and obtained consent from both the children and their parents. The study followed a structured procedure: (1) Collecting children's resting heart rate for baseline ECG data collection, followed by completing the pretest questionnaire to measure hope and travel expectations. (2) An 11-minute VR experience while ECG data were recorded. (3) Completing the posttest questionnaire to measure hope, travel expectations, and pleasure outcomes.

The experiment was conducted individually in classrooms (healthy children) or hospital wards (hospitalized children). .

Data Analysis

All scales were tested using SPSS software (version 20; IBM Corp). We assessed the reliability of the items by examining the Cronbach alpha values, all of which exceeded the common threshold of 0.70, suggesting adequate reliability [41]. The pleasure scale demonstrated strong internal consistency, with a Cronbach α value of 0.88. Similarly, the hope scale exhibited high reliability, with a Cronbach α value of 0.93. Additionally, the travel expectations scale showed adequate internal consistency, with a Cronbach α value of 0.73. To analyze the data related to hope, travel expectations, and pleasure—key factors for a successful VR intervention—the questionnaire findings were compared between groups using an independent

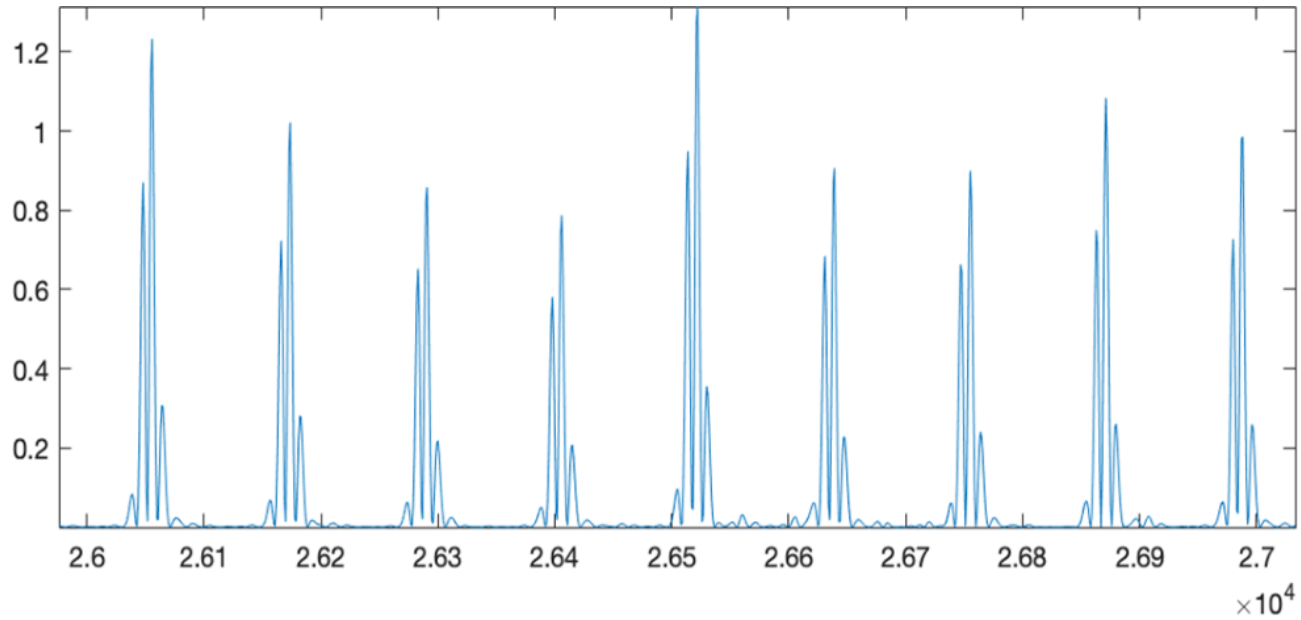
XSL•FO
RenderX

2-tailed *t* test. Additionally, post hoc analyses were conducted using MANOVA to examine differences between children who had traveled (once or more) and those who had not. Independent samples 2-tailed *t* tests were also performed to compare children with and with no prior VR experience.

An ECG typically requires a physiological signal length of 30 seconds to 5 minutes for heartbeat detection [42]; therefore, this study used 5 minutes as the benchmark. To extract the R-peak of the QRS complex (the QRS complex is the portion of the ECG that reflects the depolarization of the ventricles, marking the onset of ventricular contraction; it is composed of the Q, R, and S waves and is a key indicator of cardiac electrical activity), we applied the Pan-Tompkins open-source algorithm using MATLAB [43]. During the process of collecting physiological signals, the original signal is prone to noise interference, making the QRS complex of the ECG difficult to distinguish. To address this, a filter was used to remove unnecessary signals, enhancing the ECG waveform's characteristics.

Since heart rate variability is calculated based on the highest peak (R-peak) in the ECG, a Butterworth high-pass filter was applied to eliminate unwanted frequency bands and highlight the R-peak within the QRS complex. To detect the position of the R-peak, the differentiated signal was first segmented. Then, squaring was applied to amplify the signal, making high-frequency values more prominent, as shown in Figure 2.

Figure 2. Actual physiological signal after square.



Finally, the moving window integral method was used, with 30 samples (approximately 150 ms) as the window size. When both the difference threshold and the integration threshold met the required conditions simultaneously, an R-peak was detected, and the threshold was automatically updated.

As shown in Figure 3, the black line represents the noise threshold curve, the green line represents the automatically adapted threshold curve, and the red line represents the predicted R-peak curve. Finally, the results of the R-peak sequence were used to plot the pulse sequence onto the original waveform, as shown in Figure 4, which can be used to evaluate accuracy.

XSL•FO
RenderX

Figure 3. Actual physiological signal results after moving window integral.

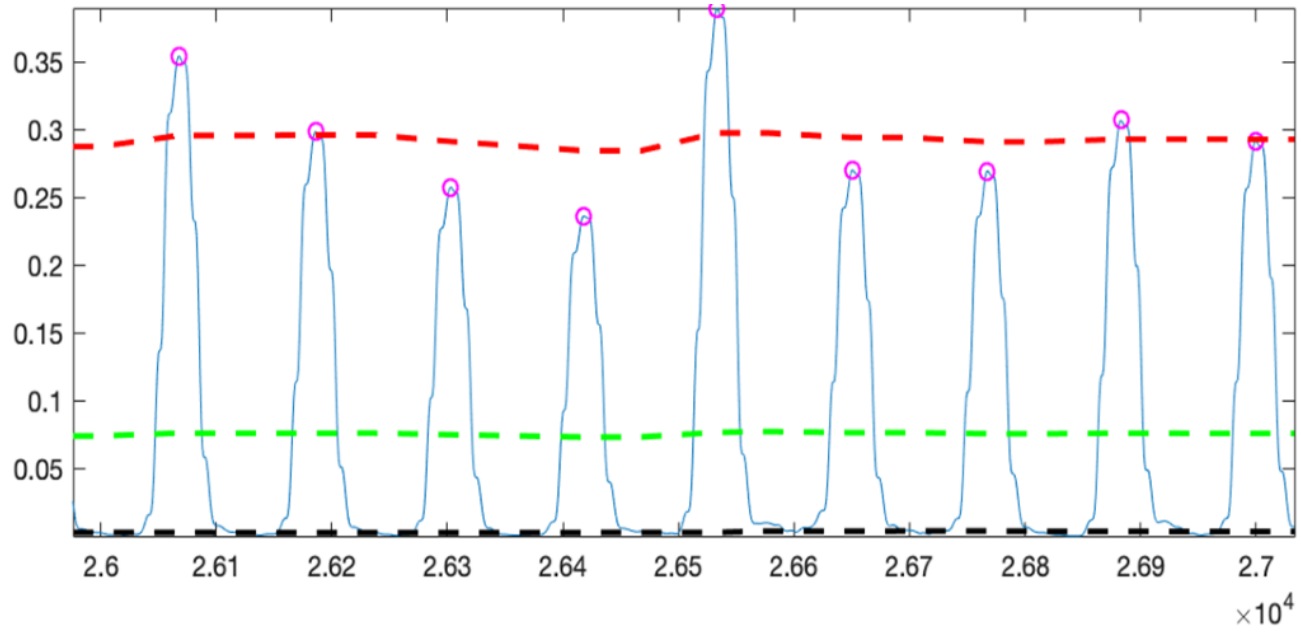
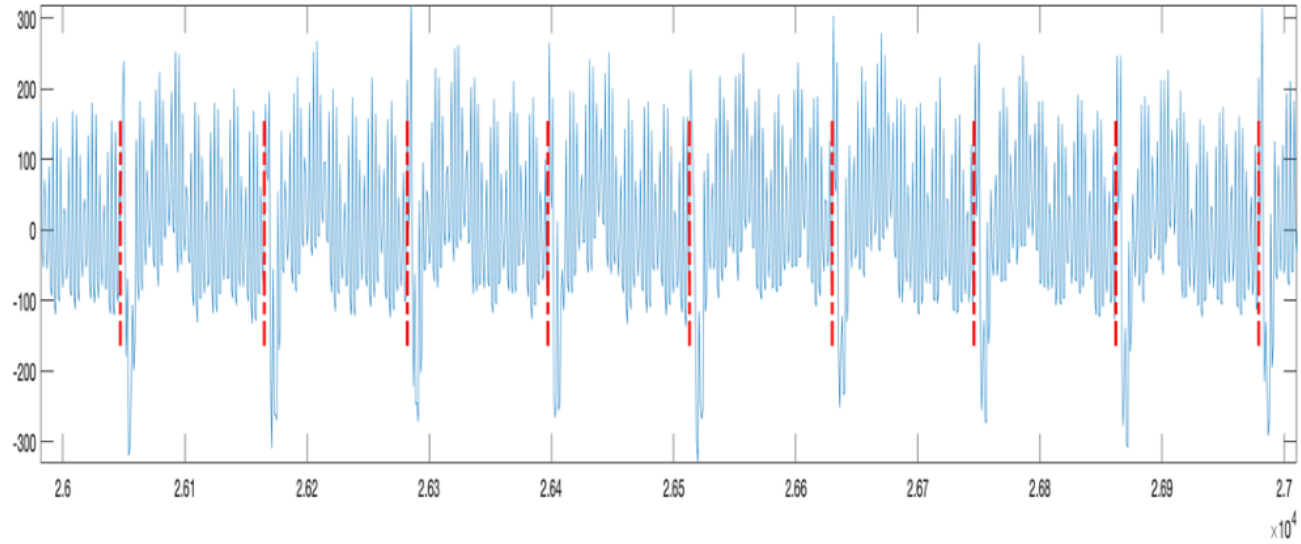


Figure 4. R-peak pulse sequence diagram.



Results

Sample Characteristics

The participants included 18 hospitalized children and 30 healthy children, all aged between 7 and 18 years (see Table 2 for details). The hospitalized children were preselected by the attending physician and were patients with leukemia with stable health who were willing to participate in the experiment.

XSL•FO
RenderX

Table . Descriptive statistics of participants in the formal test by group.

	Hospitalized children (n=18) , n (%)	Healthy children (n=30), n (%)
Gender		
Boy	11 (61)	23 (77)
Girl	7 (39)	7 (23)
Age (years)		
7 - 9	6 (33)	N/A[a]
10 - 12	3 (17)	18 (60)
13 - 15	6 (33)	12 (40)
16 - 18	3 (17)	N/A
Experience of traveling abroad		
0 times	8 (45)	11 (37)
1 - 5 times	6 (33)	17 (57)
5 - 10 or >10 times	4 (22)	2 (6)
Experience of using VR[b]		
Used	8 (44)	9 (30)
Never used	10 (56)	21 (70)

[a]N/A: not applicable.

[b]VR: virtual reality.

We use the MANOVA test for post hoc analyses to compare 2 different dependent variables (hope and travel expectations) among 3 groups based on their experience of traveling abroad (0 times, 1 - 5 times, and more than 10 times) in the hospitalized children. The output shows that there are no significant differences in hope among groups (experience of traveling abroad 0 times: mean 5.34, SD 1.15; 1 - 5 times: mean 5.44, SD 1.11; P=.99; and 0 times: mean 5.34, SD 1.15; more than 10 times: mean 5.34, SD 1.40; P>.99). Similarly, there are no significant differences in travel expectations among groups (experience of traveling abroad 0 times: mean 6.14, SD 0.85; 1 - 5 times: mean 5.75, SD 1.00; P=.70; and 0 times: mean 6.14, SD 0.85; 5 - 10 times: mean 6.31, SD 0.94; P=.94). Also, the results show that there are no significant differences in hope among groups in the healthy children.

In addition, we adopt an independent 2-tailed t test to analyze 2 different dependent variables (hope and travel expectations) between 2 groups based on their experience of using VR (used and never used) in hospitalized and healthy children, respectively. In the hospitalized children group, the results show that there are no significant differences in hope between the 2 groups (experience of using VR: used vs never used; t=0.57; P=.58), and no significant differences in travel expectations between the 2 groups (t_{17}=−0.04; P=.97). Similarly, the results show that there are no significant differences in hope and travel expectations in the healthy children group.

Hope, Pleasure, and Travel Expectations Results

First, there was no significant difference in the sense of hope between the hospitalized and healthy children before using VR (t_{46}=−1.02; P=.32). However, after using VR to experience tourist attractions, a significant difference was observed between the 2 groups (t_{46}=−2.06; P=.049) (Table 3).

XSL•FO
RenderX

Table . Levels of hope before and after watching virtual reality in the hospitalized and healthy children during the formal test[a].

Stage and group	Mean (SD)	*t* test (*df*)	*P* value[b]
Before watching VR[c]		−1.02 (46)	.32
Hospitalized children	5.51 (1.17)		
Healthy children	5.83 (0.87)		
After watching VR		−2.06 (46)	.049
Hospitalized children	5.73 (1.15)		
Healthy children	6.36 (0.76)		

[a]An independent samples 2-tailed *t* test was conducted to compare levels of hope between the hospitalized children and the healthy children before and after watching VR.

[b]Significance at .05 level.

[c]VR: virtual reality.

To further examine the degree of improvement in hope after the experiment, we conducted paired samples 2-tailed *t* tests for each group separately. The results showed significant differences in both the hospitalized children group (t_{17}=−2.46; *P*=.03) and the healthy children group (t_{29}=−3.46; *P*=.002). Furthermore, an independent samples 2-tailed *t* test revealed a significant difference in postintervention hope levels between the hospitalized and healthy children (t_{46}=−2.06; *P*=.049). Then, we further compared the mean differences in hope scores from preintervention to postintervention using an independent samples 2-tailed *t* test and found that the difference was statistically significant (t_{46}=2.18; *P*=.03). Healthy children demonstrated a greater increase in mean hope scores (an increase of 0.53, from 5.83 to 6.36) compared with the hospitalized children (an increase of 0.22, from 5.51 to 5.73). This suggest that using VR to experience tourist attractions may be more effective in enhancing the sense of hope in healthy children (Table 4). On the other hand, this result may also indicate that healthy children's sense of hope is more easily stimulated than that of hospitalized children.

Table . Comparison of hope levels before and after watching virtual reality within the hospitalized and healthy children during the formal test[a].

Group and stage		Mean (SD)	*t* test (*df*)	*P* value[b]
Hospitalized children			−2.46 (17)	.03
	Before watching VR[c]	5.51 (1.17)		
	After watching VR	5.73 (1.15)		
Healthy children			−3.46 (29)	.002
	Before watching VR	5.83 (0.865)		
	After watching VR	6.36 (0.756)		

[a]Paired samples 2-tailed *t* tests were conducted separately within the hospitalized children and healthy children to examine the change in hope levels before and after watching VR.

[b]Significance at .05 level.

[c]VR: virtual reality.

In addition, the results of the independent samples 2-tailed *t* test revealed a nonsignificant difference in pleasure (t_{46}=−1.116; *P*=.27) and travel expectations (t_{46}=−0.300; *P*=.77) between the hospitalized and healthy children. We inferred that both hospitalized and healthy children experienced pleasure when using VR, likely because it was a novel experience for them, and they had expectations about traveling.

Power and Effect Size for Independent Samples 2-Tailed *t* Test

For the independent samples 2-tailed *t* test, Cohen *d* is calculated by taking the mean difference between the 2 groups and dividing it by the pooled SD. Cohen *d* is an appropriate effect size measure when both groups have similar SDs and equal sample sizes. However, since our sample sizes differ, we used Hedges *g*, which adjusts for sample size differences. In this study, Hedges *g* was calculated as:

Hedges g=(6.4−5.7)0.967291=0.7

According to Zach [44], a *g* value of 0.2 represents a small effect size, 0.5 represents a medium effect size, and 0.8 represents a large effect size. Therefore, the effect size in this study (0.7) is close to a large effect size.

ECG Data Analysis Result

The resting state represents the baseline heartbeat. In this study, it was used to determine whether the heart rates of the 2 groups (hospitalized children and healthy children) were at the same level before the experiment. The results showed no significant difference between the groups before using VR.

XSL•FO
RenderX

However, when using VR, 5 indicators showed significant differences between the 2 groups: mean normal-to-normal intervals (t_{46}=−2.119; P=.04), normal-to-normal intervals (t_{46}=−2.856; P=.007), SD of normal-to-normal intervals (SDNN; t_{46}=2.189; P=.04), low-frequency (LF; t_{46}=2.834; P=.008), and high-frequency normalized units (HF_nu;t_{46}=2.297; P=.03). The results are summarized in Table 5.

Table . Differences in electrocardiography indicators between the hospitalized and healthy children while watching virtual reality[a].

Autonomic nervous system, indicator, and group		Mean (SD)	*t* test (*df*)	*P* value[b]
Parasympathetic nerves				
SDNN[c]			2.19 (46)	.04
	Group 1[d]	150.08 (43.75)		
	Group 2[e]	119.12 (38.59)		
HF_nu[f]			2.30 (46)	.03
	Group 1	74.82 (15.32)		
	Group 2	64.38 (11.03)		
Sympathetic nerve				
LF[g]			−2.83 (46)	.008
	Group 1	192.77 (220.48)		
	Group 2	443.64 (307.65)		

[a]This table shows the results of independent samples 2-tailed *t* tests comparing electrocardiography indicators between hospitalized children and healthy children during virtual reality viewing.

[b]Significance at .05 level.

[c]SDNN: SD of all normal-to-normal intervals.

[d]Group 1: hospitalized children.

[e]Group 2: healthy children.

[f]HF_nu: high-frequency normalized units.

[g]LF: low-frequency power.

In the ECG analysis, index calculations can be performed after obtaining the R-R interval. According to the literature, the autonomic nervous system (ANS) is mainly composed of the sympathetic nervous system (SNS) and the parasympathetic nervous system (PNS). If the ANS is biased toward the SNS, the PNS has a smaller effect. To illustrate, if the ANS were compared with a car, the SNS would represent the accelerator, and the PNS would represent the brake.

ECG indicators, such as SDNN and HF_nu, are related to the PNS, while LF is related to the SNS. In this study, the mean value of the SNS indicator (LF) in the hospitalized children (mean 192.77, SD 220.48) was significantly lower than the mean value of the SNS indicator (LF) in the healthy children (mean 443.64, SD 307.65). Conversely, the mean values of parasympathetic nervous indicators (SDNN, HF_nu) were significantly higher in the hospitalized children than those in the healthy children. Based on the significant results from LF, it appears that the SNS is highly active during VR exposure, indicating that healthy children experience VR in a more excited state. The results of the formal test are summarized in Table 6.

XSL•FO
RenderX

Table . Summary of formal test results comparing the hospitalized and healthy children across physiological and psychological measures.

Measurement and results		Significance
ECG[a]		
	Comparison of the ECG indicators of G1[b] and G2[c] when in a resting state.	Not significant
	Comparison of the ECG indicators of G1 and G2 when using VR[d].	Significant
Hope		
	Comparison of the state of hope in G1 and G2 before using VR.	Not significant (*P*=.32)
	Comparison of the state of hope in G1 and G2 after using VR.	Significant (*P*=.05)
	Improvement of G1's state of hope.	Significant (*P*=.03); the mean value rose by 0.22, from 5.51 to 5.73.
	Improvement of G2's state of hope.	Significant (*P*=.002); the mean value rose by 0.53, from 5.83 to 6.36.
Pleasure		Not significant (*P*=.27)
Travel expectations		Not significant (*P*=.77)

[a]ECG: electrocardiography.

[b]G1: hospitalized children with leukemia.

[c]G2: healthy children.

[d]VR: virtual reality.

Discussion

Summary of Key Findings

We assessed the effects of VR on hope, pleasure, and travel expectations in healthy and hospitalized children using ECG and questionnaires. The study demonstrated consistency between self-reported data and physiological measures, suggesting that VR effectively enhanced hope among both healthy and hospitalized children with leukemia, with a more pronounced effect in the healthy group. However, no significant changes were observed in pleasure or travel expectations in either group.

Clinical and Research Implications

While questionnaires captured subjective experiences, ECG objectively measured real-time physiological responses. Indicators such as SDNN and LF revealed differences in ANS activity. Healthy children exhibited higher sympathetic activation, reflecting greater engagement, while hospitalized children showed more subdued responses. This alignment validates the questionnaire measures and enhances the study's reliability. Beyond validation, ECG plays a crucial role in detecting subtle or unconscious emotional responses and tracking real-time fluctuations during VR exposure. By linking subjective perception with objective data, ECG strengthens methodological rigor and supports future VR-based emotional and physiological research.

We infer that this result possibly reflects hospitalized children's deeper psychological needs and the greater impact of long-term hospitalization stress. Their sense of hope may be more focused on overcoming illness and achieving milestone progress in treatment. It is likely linked to improvements in their health condition, disease remission, or the possibility of participating in activities they have missed due to their illness. Therefore, this suggests that when designing interventions for hospitalized children, more targeted emotional support strategies should be considered. This aligns with previous research, in which Napetschnig et al [45] proposed that promoting user motivation is one of the key quality criteria for developing user-centered VR.

Nevertheless, we found a significant difference in the mean hope scores before and after using VR, both in healthy children and in hospitalized children. VR is a promising technique for cognitive diversion and psychological rehabilitation support for patients with cancer [46], particularly when they experience boredom during hospitalization following treatment [47,48]. Such computerized technologies are especially appealing to this age group, which, in turn, increases their enthusiasm [46]. For example, the results revealed no significant difference in pleasure and travel expectations between the hospitalized and healthy children. We infer that this finding suggests that both hospitalized and healthy children experienced pleasure when using VR because it was a novel experience for them and sparked their expectations about traveling. Although some participants in the hospitalized children group had prior experience using VR, there were no significant differences in hope and travel expectations between those who had used VR before and those who had not. Therefore, we suggest that health care workers shift their focus from health to overall well-being [49] and incorporate immersive VR experiences into pediatric care as a form of distraction therapy. This approach can help children cope with long-term hospitalization and enhance their emotional well-being.

XSL•FO
RenderX

Study Limitations and Recommendations for Future Research

Some limitations of this study should be noted. First, the study examined differences between healthy children and hospitalized children with leukemia aged 7 - 18 years. Therefore, the results are applicable only to hospitalized children with leukemia in the midstage of treatment with stable vital signs and not to those with different medical conditions or from other age groups. Future studies should investigate the impact of VR apps on other age groups and medical conditions. Second, the VR experience content used in this research was primarily based on tourist attractions. Future studies could explore different VR content and experiences. Third, we acknowledge the limitations of ECG in differentiating specific emotional responses, as an increased heart rate can be associated with multiple emotions such as joy, surprise, or satisfaction. This is why we combined questionnaires to enhance the interpretation of emotional responses. Future studies could use more precise psychophysiological methods, such as functional magnetic resonance imaging, to improve the accuracy of emotional response assessments.

Conclusions

With the advancement of emerging technologies, this study aimed to evaluate the pleasurable experience of VR and its impact on hospitalized children's travel expectations and hope by using ECG and questionnaires. We examined the differences between healthy and hospitalized children in their perceptions of pleasure and hope to provide future scholars with a reference for research on hospitalized children. In this study, VR provides hospitalized children with a safe space and a more immersive way to explore the world. It also provides another way for children to experience the world in addition to books and photographs, and it gives children more opportunities to learn about the world. Through self-learning and adaptation, children can find ways to deal with their emotions and relieve their psychological pressure. VR apps are considered an accessible, low-risk, and cost-effective innovative approach for children with leukemia, particularly during repeated waves of the epidemic, as they minimize the risk of cross-infection [50].

Acknowledgments

The manuscript's English-language quality was enhanced using artificial intelligence–assisted proofreading tools, while maintaining the original content and meaning. The authors would like to thank the participants who joined the virtual reality experiment and the 2 research assistants who helped recording participants' electrocardiography data in the study. The authors also deeply appreciate Fang-Liang Huang, the director of Children's Medical Center, Taichung Veterans General Hospital, who provided valuable perspectives of the virtual reality experiment. This study was supported by the grants from Taichung Veterans General Hospital and Tunghai University (TCVGH-T1107809).

Data Availability

The data associated with this article are available from the corresponding author upon reasonable request.

Authors' Contributions

PSH and HSC conceptualized the study. FLH and WHC contributed to data collection. PSH, HSC, and WHC conducted data analysis and interpretation. PSH and WHC wrote the manuscript draft and revised the manuscript. PSH, FLH, HSC, and WHC contributed to the final approval of the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Ahmadpour N, Weatherall AD, Menezes M, Yoo S, Hong H, Wong G. Synthesizing multiple stakeholder perspectives on using virtual reality to improve the periprocedural experience in children and adolescents: survey study. J Med Internet Res 2020 Jul 17;22(7):e19752. [doi: 10.2196/19752] [Medline: 32706671]
2. Scherz WD, Magaña VC, Seepold R, Madrid NM, Ortega JA. Can virtual reality be used as a significant stressor for studies using ECG? Procedia Comput Sci 2020;176:3255-3262. [doi: 10.1016/j.procs.2020.09.123]
3. Bun P, Gorski F, Grajewski D, Wichniarek R, Zawadzki P. Low-cost devices used in virtual reality exposure therapy. Procedia Comput Sci 2017;104:445-451. [doi: 10.1016/j.procs.2017.01.158]
4. Pourmand A, Davis S, Lee D, Barber S, Sikka N. Emerging utility of virtual reality as a multidisciplinary tool in clinical medicine. Games Health J 2017 Oct;6(5):263-270. [doi: 10.1089/g4h.2017.0046] [Medline: 28759254]
5. Levac D, Glegg S, Colquhoun H, Miller P, Noubary F. Virtual reality and active videogame-based practice, learning needs, and preferences: a cross-Canada survey of physical therapists and occupational therapists. Games Health J 2017 Aug;6(4):217-228. [doi: 10.1089/g4h.2016.0089] [Medline: 28816511]

XSL•FO
RenderX

6. Jiang N, Zhang Y, Liang S, et al. Effectiveness of virtual simulations versus mannequins and real persons in medical and nursing education: meta-analysis and trial sequential analysis of randomized controlled trials. J Med Internet Res 2024 Dec 5;26:e56195. [doi: 10.2196/56195] [Medline: 39636688]
7. Smolarski K, Leone K, Robbins SJ. Reducing negative mood through drawing: comparing venting, positive expression, and tracing. Art Ther (Alex) 2015 Oct 2;32(4):197-201. [doi: 10.1080/07421656.2015.1092697]
8. Giebel GD, Raszke P, Nowak H, et al. Problems and barriers related to the use of AI-based clinical decision support systems: interview study. J Med Internet Res 2025 Feb 3;27:e63377. [doi: 10.2196/63377] [Medline: 39899342]
9. Jans C, Bogossian F, Andersen P, Levett-Jones T. Examining the impact of virtual reality on clinical decision making—an integrative review. Nurse Educ Today 2023 Jun;125(105767):105767. [doi: 10.1016/j.nedt.2023.105767] [Medline: 36906980]
10. Lan L, Mao RQ, Qiu RY, Kay J, de Sa D. Immersive virtual reality for patient-specific preoperative planning: a systematic review. Surg Innov 2023 Feb;30(1):109-122. [doi: 10.1177/15533506221143235] [Medline: 36448920]
11. Zheng Y, Chen Y, Chen Y, et al. Metaverse clinic for pregnant women with subclinical hypothyroidism: prospective randomized study. J Med Internet Res 2025 Feb 5;27:e64634. [doi: 10.2196/64634] [Medline: 39908543]
12. van der Kruk SR, Zielinski R, MacDougall H, Hughes-Barton D, Gunn KM. Virtual reality as a patient education tool in healthcare: a scoping review. Patient Educ Couns 2022 Jul;105(7):1928-1942. [doi: 10.1016/j.pec.2022.02.005]
13. Zhang C. The why, what, and how of immersive experience. IEEE Access 2020;8:90878-90888. [doi: 10.1109/ACCESS.2020.2993646]
14. Trahan MH, Smith KS, Talbot TB. Past, present, and future: editorial on virtual reality applications to human services. J Technol Hum Serv 2019 Jan 2;37(1):1-12. [doi: 10.1080/15228835.2019.1587334]
15. Bryson S. Virtual reality: a definition, history, and a personal essay. arXiv. Preprint posted online on Dec 16, 2013. [doi: 10.48550/arXiv.1312.4322]
16. Kober SE, Kurzmann J, Neuper C. Cortical correlate of spatial presence in 2D and 3D interactive virtual reality: an EEG study. Int J Psychophysiol 2012 Mar;83(3):365-374. [doi: 10.1016/j.ijpsycho.2011.12.003] [Medline: 22206906]
17. Maschio T. Storyliving: an ethnographic study of how audiences experience VR and what that means for journalists. Google News Lab. 2017. URL: https://newslab.withgoogle.com/assets/docs/storyliving-a-study-of-vr-in-journalism.pdf [accessed 2025-03-12]
18. Vitek KN, Yeater EA. Women's response performance in sexually risky situations: associations with disinhibited sex-related alcohol expectancies. J Interpers Violence 2022 Aug;37(15-16):NP14854-NP14876. [doi: 10.1177/08862605211016354] [Medline: 33988045]
19. Bossenbroek R, Wols A, Weerdmeester J, Lichtwarck-Aschoff A, Granic I, van Rooij MMJW. Efficacy of a virtual reality biofeedback game (DEEP) to reduce anxiety and disruptive classroom behavior: single-case study. JMIR Ment Health 2020 Mar 24;7(3):e16066. [doi: 10.2196/16066] [Medline: 32207697]
20. Hadley W, Houck C, Brown LK, Spitalnick JS, Ferrer M, Barker D. Moving beyond role-play: evaluating the use of virtual reality to teach emotion regulation for the prevention of adolescent risk behavior within a randomized pilot trial. J Pediatr Psychol 2019 May 1;44(4):425-435. [doi: 10.1093/jpepsy/jsy092] [Medline: 30551157]
21. Evans E, Naugle KE, Kaleth AS, Arnold B, Naugle KM. Physical activity intensity, perceived exertion, and enjoyment during head-mounted display virtual reality games. Games Health J 2021 Oct;10(5):314-320. [doi: 10.1089/g4h.2021.0036] [Medline: 34449262]
22. Ridout B, Kelson J, Campbell A, Steinbeck K. Effectiveness of virtual reality interventions for adolescent patients in hospital settings: systematic review. J Med Internet Res 2021 Jun 28;23(6):e24967. [doi: 10.2196/24967] [Medline: 34185015]
23. Hammesfahr L. Sick kids can now see the world without leaving St. Jude Hospital. Scripps News. 2024. URL: https://www.scrippsnews.com/us-news/sick-kids-can-now-see-the-world-without-leaving-st-jude-hospital [accessed 2025-03-10]
24. Mitsea E, Drigas A, Skianis C. Well-being technologies and positive psychology strategies for training metacognition, emotional intelligence and motivation meta-skills in clinical populations: a systematic review. Psych 2024;6(1):305-344. [doi: 10.3390/psych6010019]
25. van Gelderen MJ, Nijdam MJ, Haagen JFG, Vermetten E. Interactive motion-assisted exposure therapy for veterans with treatment-resistant posttraumatic stress disorder: a randomized controlled trial. Psychother Psychosom 2020;89(4):215-227. [doi: 10.1159/000505977] [Medline: 32203971]
26. Bush NE, Smolenski DJ, Denneson LM, Williams HB, Thomas EK, Dobscha SK. A virtual hope box: randomized controlled trial of a smartphone app for emotional regulation and coping with distress. Psychiatr Serv 2017 Apr 1;68(4):330-336. [doi: 10.1176/appi.ps.201600283] [Medline: 27842473]
27. Mabrook R, Singer JB. Virtual reality, 360° video, and journalism studies: conceptual approaches to immersive technologies. J Stud 2019 Oct 26;20(14):2096-2112. [doi: 10.1080/1461670X.2019.1568203]
28. Marín-Morales J, Higuera-Trujillo JL, Greco A, et al. Affective computing in virtual reality: emotion recognition from brain and heartbeat dynamics using wearable sensors. Sci Rep 2018 Sep 12;8(1):13657. [doi: 10.1038/s41598-018-32063-4] [Medline: 30209261]
29. Preston AM, Brown L, Padala KP, Padala PR. Veterans Affairs health care provider perceptions of virtual reality: brief exploratory survey. Interact J Med Res 2022;11(2):e38490. [doi: 10.2196/38490]

XSL•FO
RenderX

30. Bakker A, Janssen L, Noordam C. Home to hospital live streaming with virtual reality goggles: a qualitative study exploring the experiences of hospitalized children. JMIR Pediatr Parent 2018 Dec 13;1(2):e10. [doi: 10.2196/pediatrics.9576] [Medline: 31518293]
31. Dascal J, Reid M, IsHak WW, et al. Virtual reality and medical inpatients: a systematic review of randomized, controlled trials. Innov Clin Neurosci 2017;14(1-2):14-21. [Medline: 28386517]
32. Zhang W, Shu L, Xu X, Liao D. Affective virtual reality system (AVRS): design and ratings of affective VR scenes. 2017 Presented at: 2017 International Conference on Virtual Reality and Visualization (ICVRV); Oct 21-22, 2017; Zhengzhou, China p. 311-314. [doi: 10.1109/ICVRV.2017.00072]
33. Schmidt P, Reiss A, Duerichen R, Marberger C, Van Laerhoven K. Introducing WESAD, a multimodal dataset for wearable stress and affect detection. Presented at: ICMI '18: International Conference on Multimodal Interaction; Oct 16-20, 2018; Boulder, CO p. 400-408. [doi: 10.1145/3242969.3242985]
34. Ahmad Z, Rabbani S, Zafar MR, Ishaque S, Krishnan S, Khan N. Multilevel stress assessment from ECG in a virtual reality environment using multimodal fusion. arXiv. Preprint posted online on Jul 9, 2021. [doi: 10.48550/arXiv.2107.04566]
35. von Rosenberg W, Chanwimalueang T, Adjei T, Jaffer U, Goverdovsky V, Mandic DP. Resolving ambiguities in the LF/HF ratio: LF-HF scatter plots for the categorization of mental and physical stress from HRV. Front Physiol 2017;8:360. [doi: 10.3389/fphys.2017.00360]
36. Kalatzis A, Stanley L, Prabhu VG. Affective state classification in virtual reality environments using electrocardiogram and respiration signals. In: 2021 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR): IEEE; 2021:160-167. [doi: 10.1109/AIVR52153.2021.00037]
37. Bakhuis W, Max SA, Maat A, Bogers A, Mahtab EAF, Sadeghi AH. Preparing for the future of cardiothoracic surgery with virtual reality simulation and surgical planning: a narrative review. Shanghai Chest 2023;7:23. [doi: 10.21037/shc-22-63]
38. Schifferstein HNJ, Tanudjaja I. Visualising fragrances through colours: the mediating role of emotions. Perception 2004;33(10):1249-1266. [doi: 10.1068/p5132] [Medline: 15693669]
39. Snyder CR, Hoza B, Pelham WE, et al. The development and validation of the Children's Hope Scale. J Pediatr Psychol 1997 Jun;22(3):399-421. [doi: 10.1093/jpepsy/22.3.399] [Medline: 9212556]
40. Wong MMC, Cheung R, Wan C. A study on traveler expectation, motivation and attitude. Contemp Manag Res 2013 Jun 30;9(2):169-186. [doi: 10.7903/cmr.11023]
41. Nunnally JC. Psychometric Theory: McGraw-Hill; 1978.
42. Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Comput 2006 Dec;44(12):1031-1051. [doi: 10.1007/s11517-006-0119-0]
43. Sedghamiz H. Matlab Implementation of Pan Tompkins ECG QRS Detector: DataCite; 2014.
44. Zach B. What is Hedges' g? (definition & example). Statology. 2021. URL: https://www.statology.org/hedges-g.Statology [accessed 2022-10-15]
45. Napetschnig A, Brixius K, Deiters W. Development of a core set of quality criteria for virtual reality applications designed for older adults: multistep qualitative study. Interact J Med Res 2023 Sep 27;12(1):e45433. [doi: 10.2196/45433] [Medline: 37756112]
46. Shiri S, Tenenbaum A, Sapir-Budnero O, Wexler ID. Elevating hope among children with attention deficit and hyperactivity disorder through virtual reality. Front Hum Neurosci 2014;8:198. [doi: 10.3389/fnhum.2014.00198] [Medline: 24847233]
47. Rutkowski S, Czech O, Wrzeciono A, Kiper P, Szczepańska-Gieracha J, Malicka I. Virtual reality as a chemotherapy support in treatment of anxiety and fatigue in patients with cancer: a systematic review and meta-analysis and future research directions. Complement Ther Med 2021 Sep;61:102767. [doi: 10.1016/j.ctim.2021.102767] [Medline: 34403772]
48. Zasadzka E, Pieczyńska A, Trzmiel T, Hojan K. Virtual reality as a promising tool supporting oncological treatment in breast cancer. Int J Environ Res Public Health 2021 Aug 19;18(16):8768. [doi: 10.3390/ijerph18168768] [Medline: 34444513]
49. Smits M, Ludden GDS, Verbeek PP, van Goor H. How digital therapeutics are urging the need for a paradigm shift: from evidence-based health care to evidence-based well-being. Interact J Med Res 2022 Oct 20;11(2):e39323. [doi: 10.2196/39323] [Medline: 36264624]
50. Horton R. Offline: the case for No-COVID. Lancet 2021 Jan;397(10272):359. [doi: 10.1016/S0140-6736(21)00186-0]

Abbreviations

ANS: autonomic nervous system
ECG: electrocardiography
HF_nu: high-frequency normalized unit
LF: low-frequency power
PNS: parasympathetic nervous system
SDNN: SD of all normal-to-normal intervals
SNS: sympathetic nervous system
VR: virtual reality

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 12.08.24; peer-reviewed by DA Jelincic, D Poddighe, E Mitsea, OR Runswick; revised version received 29.03.25; accepted 31.03.25; published 16.06.25.

Please cite as:
Hsieh PS, Chiang HS, Huang FL, Chi WH
The Effects of Virtual Reality on Hope and Travel Expectations in Healthy and Hospitalized Children: Quasi-Experimental Design Approach
Interact J Med Res 2025;14:e65311
URL: https://www.i-jmr.org/2025/1/e65311
doi:10.2196/65311

© Pei-Shan Hsieh, Hsiu-Sen Chiang, Fang-Liang Huang, Wen-Hsin Chi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 16.6.2025. 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 Hot Spots and Global Research Trends in Exergaming Between 1997 and 2024: Bibliometric Analysis

Abdullah Kayhan[1], MSc; Elif Kose[1], PhD; Burcu Kayhan[2], MSc; Nan Zeng[3], PhD

[1]Department of Recreation, Faculty of Sport Sciences, Akdeniz University, Antalya, Turkey

[2]Department of Recreation, Institute of Social Sciences, Akdeniz University, Antalya, Turkey

[3]Department of Exercise and Health Sciences, Manning College of Nursing and Health Sciences, University of Massachusetts Boston, 100 Morrissey Boulevard, Boston, MA, United States

Corresponding Author:
Nan Zeng, PhD
Department of Exercise and Health Sciences, Manning College of Nursing and Health Sciences, University of Massachusetts Boston, 100 Morrissey Boulevard, Boston, MA, United States

Abstract

Background: Exergaming, the combination of exercise and gaming, has emerged as an important area in physical activity (PA) research. By leveraging advances in video game technology, exergaming supports both physical and mental health. This growing interest in exergaming has increasingly attracted researchers over the years. Examining the development of exergaming research with a bibliometric approach is thought to offer valuable perspectives to researchers by revealing the trends and main contributions in the field.

Objective: This study aims to identify the most researched concepts and topics in the field of exergaming; track the changes of trending topics over the years; identify the most influential journals as well as the authors who have contributed the most to the field; identify groundbreaking studies and neglected topics that shape future work; and reveal the countries, institutions, and collaborations that have contributed the most to the field. It also aims to identify research gaps in the field of exergaming and provide important recommendations for future research.

Methods: A bibliometric analysis covering studies between 1997 and 2024 was conducted using the Web of Science database. The R-based Bibliometrix package and the Biblioshiny web interface were used for data analysis and visualization. The analysis included original research papers and reviews. These analyses provided insights into research trends, citation metrics, and thematic developments.

Results: A total of 1626 studies were analyzed, and the results indicated a steep rise in exergaming research since 2015, peaking in the years 2020 - 2021. Major high-impact journals publishing in this area include *Games for Health Journal* and *International Journal of Environmental Research and Public Health*. Researchers who have contributed significantly and enriched the knowledge base of the exergaming field included Gao Zan, Eling de Bruin, and Zeng Nan. The most cited studies were classified into 2 different clusters, namely, cluster 1 that focuses on the concepts of PA, exercise, energy expenditure, and children, while cluster 2 focuses on rehabilitation, balance, adults, and aging. Medicine, information technology, and intention are some of the emerging themes. From a research productivity perspective, there is an undisputed front-runner, the United States, but substantial contributions have definitely come from either the Swiss Federal Institute of Technology or the Karolinska Institute.

Conclusions: Despite significant growth in exergaming research over the last decade, research gaps remain, particularly in understanding how exergaming can be effectively integrated into long-term PA promotion and broader health outcomes. These gaps were identified by the absence or low representation of relevant keywords (eg, “cost-effectiveness,” “community-based intervention,” and “long-term health outcomes”) in thematic mapping and keyword trend analyses and limited citation density in these areas. Future work should explore these issues more systematically to advance the field.

(Interact J Med Res 2025;14:e66738) doi:10.2196/66738

KEYWORDS

exergaming; active video games; physical activity; health outcomes; bibliometric analysis

Introduction

Physical inactivity is a significant risk factor for noncommunicable diseases worldwide [1]. Exergaming plays an important role in preventing these diseases associated with physical inactivity and promoting healthy habits [2]. “Exergaming,” also known as active video games, are video games that incorporate physical activity (PA) into the gaming concept, requiring players to move physically [3].

XSL•FO
RenderX

Leisure-time PA can be influenced by social and environmental factors (lack of time, lack of willpower, lack of friends, unsafe living environments, adverse weather conditions, socioeconomic status, health status, physical disability, etc) [4]. In this regard, exergaming can provide indoor or home-based alternatives that make PA more enjoyable and fun, thereby increasing exercise motivation [5]. Furthermore, unlike traditional sedentary video games, exergaming provides players with an optimal level of PA and double the amount of light to moderate energy expenditure compared to sedentary video gamers [6]. Therefore, exergaming has been used to promote PA and health in healthy adults [7], different age groups [8], and those with pathological conditions [9].

Exergaming increases energy expenditure and PA, promoting an active lifestyle [10]. Exergaming helps with weight management [11] and improving mental health [12] and can prevent sedentary behaviors [13]. The American College of Sports Medicine describes exergaming as "the future of fitness," and numerous studies highlight its potential to increase PA, especially among children and adolescents [14]. Furthermore, exergaming offers innovative solutions for health and rehabilitation services [15]. A recent systematic review suggested that exergaming may be recommended for patients undergoing cancer treatment [16]. Furthermore, exergaming has been shown to be effective in the rehabilitation of Parkinson disease [17] and in reducing cancer-related fatigue in children undergoing chemotherapy for leukemia [18]. Studies evaluating older adults have shown that exergaming improves various functions such as mental health, cognitive motor skills, and balance [19]. Additionally, exergaming supports psychological health by making PA more enjoyable [20].

Although a growing number of systematic reviews and meta-analyses have examined the effects of exergaming on specific outcomes such as PA promotion in children [11], older adults or clinical populations [16], cognitive function [19], balance improvement [21], and motivation [12,22], these studies primarily provide syntheses of intervention effects or health-related outcomes. However, they do not provide an overview of the evolution, structure, and dynamics of scientific output in this field over time. In contrast, bibliometric analysis sheds light on the historical development of a research field, allowing the identification of scientific production, trends, gaps, and future directions [23,24]. It also enables researchers to assess patterns of authorship, collaboration networks, and thematic trends [25].

Despite the growing popularity of exergaming, bibliometric studies specifically mapping this field remain limited. The few existing studies have largely focused on adjacent or overlapping fields such as PA, virtual reality (VR), augmented reality (AR), or rehabilitation research [26,27]. However, exergaming encompasses a wider range of technologies and applications, including motion-sensing consoles, mobile apps, and digital fitness platforms, which have not been adequately addressed in these previous reviews. Therefore, bibliometric analysis focusing specifically on exergaming is necessary to better understand its thematic breadth, interdisciplinary contributions, and underexplored areas. This study aims to address this gap by providing a scholarly map of the exergaming literature from 1997 to 2024, offering valuable insights for researchers and practitioners interested in advancing this burgeoning field.

In line with the research objective, we have formulated our research questions (RQs) as follows:

- RQ1: What is the distribution of studies published on exergaming by year?
- RQ2: What are the most prolific journals in the field of exergaming?
- RQ3: How is the distribution of the number of publications of the authors who have done the most work in the field of exergaming according to years?
- RQ4: What are the most cited studies and trending topics in the field of exergaming?
- RQ5: What is the subject clustering of studies in the field of exergaming?
- RQ6: What are the most productive countries, institutions, authors, and their collaborations?
- RQ7: What are the past, current, and emerging trends, key contributors, and collaborative networks shaping the field of exergaming research?

Methods

Analysis Tool and Database

The bibliometric analysis was conducted in the open-access R-Based Bibliometrix package designed by Aria and Cuccurullo [28]. This has been favored due to its smooth interaction with Web of Science (WoS) and fully exploits the Keyword Plus property for a beautiful analysis of exergaming research [29].

Combining data from different databases can lead to inconsistencies in citation metrics, metadata, and affiliation information, which may negatively impact the accuracy and reliability of the analysis [30]. For this reason, in this research, we decided to use 1 database for the study, that is, the WoS. WoS offers comprehensive options for analysis due to its good citation metrics, elaborate metadata structure, and important tools like Keyword Plus, which provides an increased number of terminologies in addition to keywords provided by the authors [31].

Search Strategy

The searches included peer-reviewed original and review papers published from 1997 to 2024. According to the search string applied, the earliest accessible indexed record from WoS dated back to 1997. Therefore, 1997 was used as the initial cutoff point for inclusion. Two independent researchers (AK and EK) carried out the searches, and the retrieved data were evaluated independently on a Microsoft Excel database for relevance to the study scope. Search strategies were based on a literature review, and the search prompt was created with search strategies, such as (TI=("exergame*" OR "gamercising" OR "exergaming" OR "gamercize*" OR "active video game*")). The detailed search strategy appears in Multimedia Appendix 1. The WoS search resulted in 2663 studies. No language restrictions were applied.

As part of the data cleaning process, duplicate records were removed (n=2), and author names, affiliations, and journal titles

XSL•FO
RenderX

were manually reviewed and standardized by 2 researchers to ensure consistency and reduce misclassification. Studies without accessible abstracts or full texts, those published from 2025 onward (n=5), and gray literature (n=1030) were excluded. Gray literature refers to academic work available in journals, books, print, or electronic formats that may be used commercially and are not controlled by publishing entities [32]. Ultimately, 1626 records were included in the analysis.

Data Analysis and Visualization

Studies meeting the data processing criteria were imported into the user-friendly Biblioshiny interface via R Studio. The analysis evaluated the impact of sources by examining paper counts, h-indexes, total citation (TC) counts, and citations per publication. The h-index is a key metric for assessing the impact of an author or journal [33]. Citations play a critical role in measuring the influence of sources within a research field. Network metrics, such as modularity (Q) and network density, were also used to evaluate the frequency of relationships and clustering structures. A modularity value between 0.4 and 0.8 indicates a well-structured clustering network [34]. Finally, social network analysis was used to visualize the actors and relationships in exergaming research.

Ethical Considerations

Ethics approval was not required for this study, as human participants were not involved.

Results

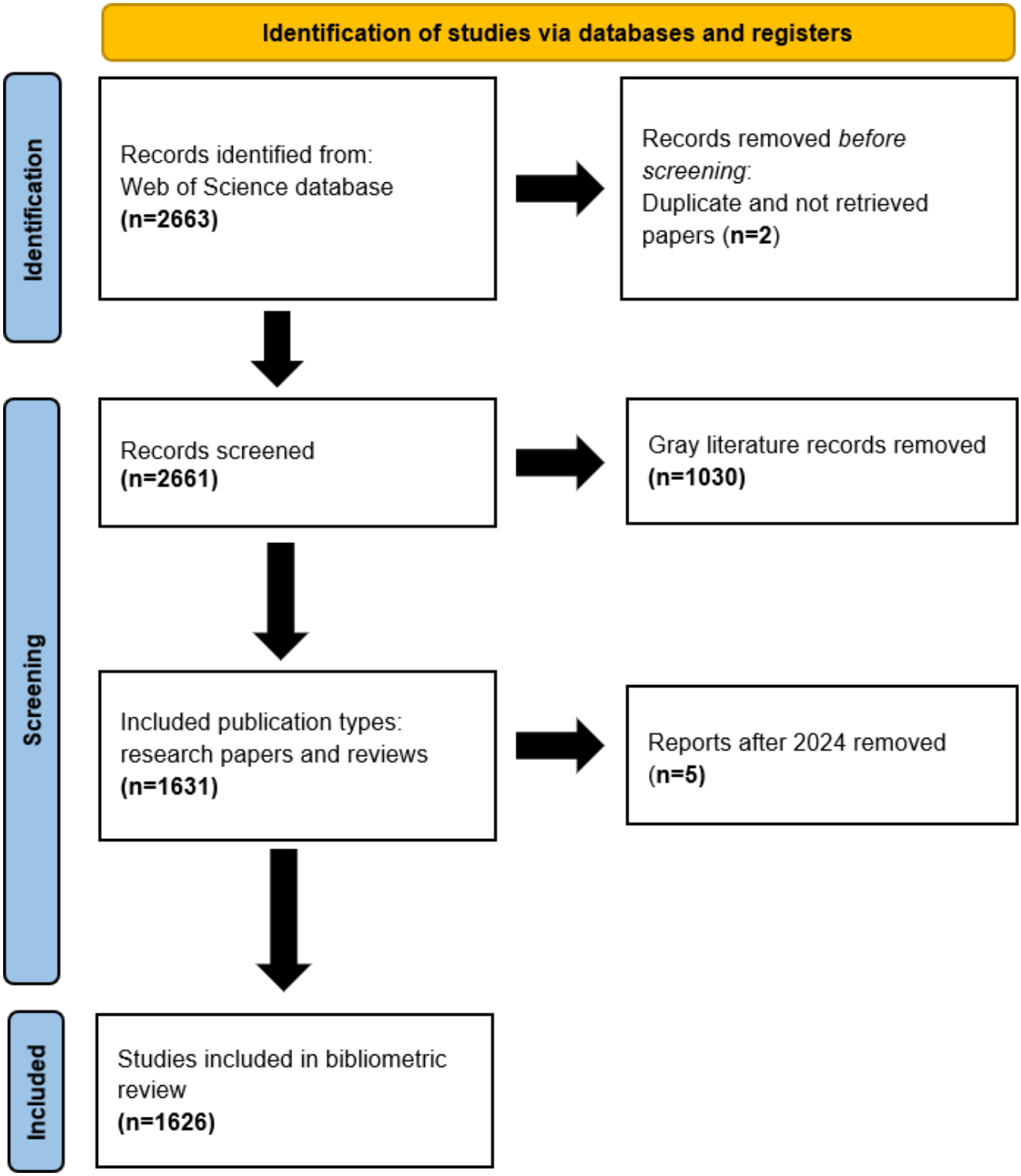
Descriptive Statistics in the Field of Exergaming

The data for the 1626 included studies are presented in Multimedia Appendix 2 and are outlined in the data collection flow diagram (Figure 1).

Table 1 provides information on the key descriptive statistics regarding studies conducted in the exergaming field. The research in this area comprises 1626 documents sourced from 553 different outlets between 1997 and 2024. The annual growth rate of publications is 22.07%, with an average document age of 5.7 years. Notably, each study has received an impressive average citation rate of 21.95%. While there are 38 single-authored documents, 6556 authors have contributed to the field, predominantly through papers (n=1402) and reviews (n=224). The rate of international author collaboration is 26.88%, with an average of 5.57 authors per document. Additionally, 2917 keywords provided by the authors enhance the richness of the dataset. Finally, the 1626 documents have been cited 45,050 times in studies indexed on WoS.

XSL•FO
RenderX

Figure 1. The flowchart for data collection.



XSL•FO
RenderX

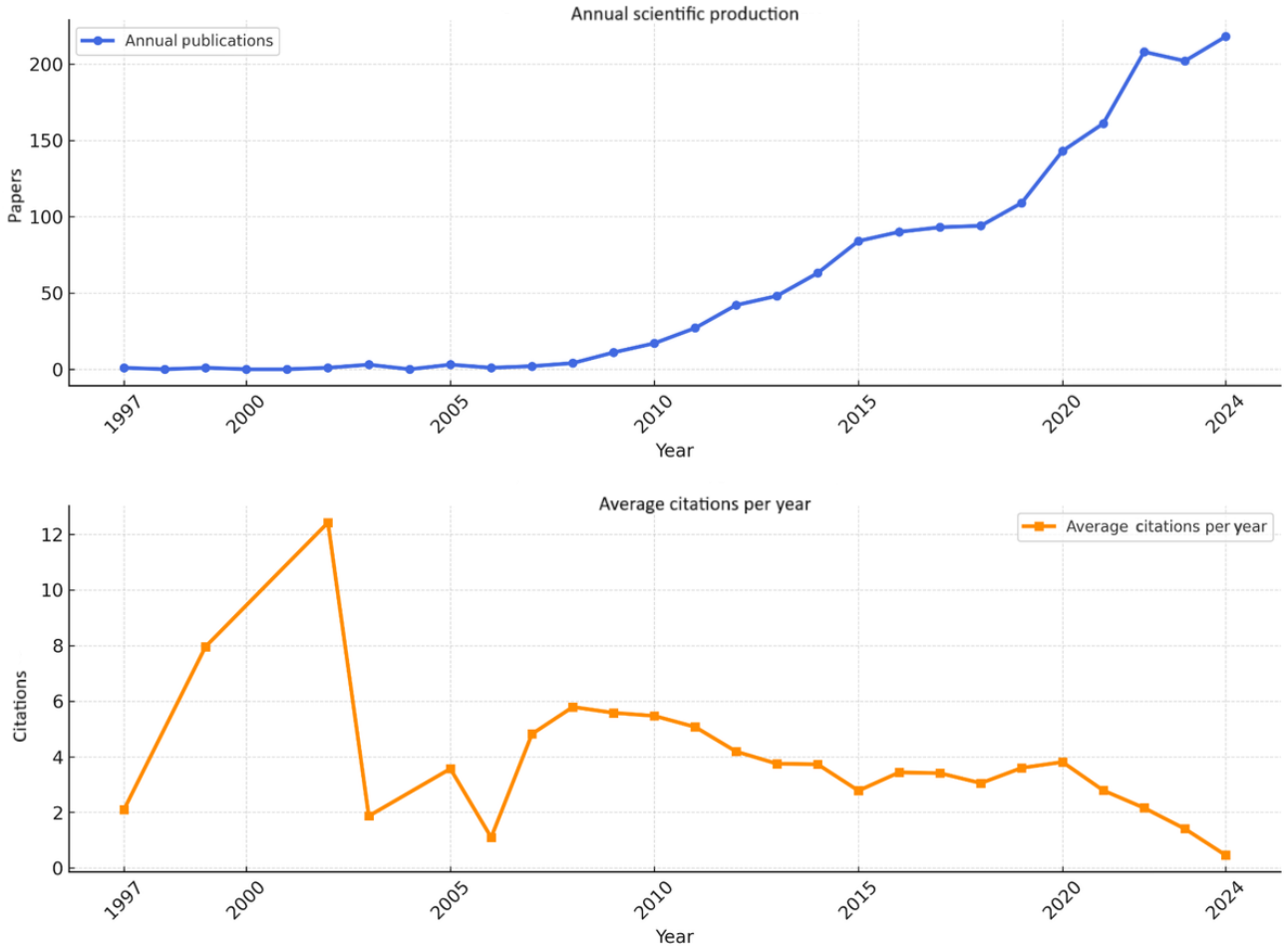
Table . Descriptive statistics of exergaming studies.

Information	Values
Timespan (years)	1997 - 2024
Sources (frequency distribution), n	553
Documents, n	1626
Annual growth rate (%)	22.07
Authors of multiauthored documents, n (%)	6556 (99.4)
Authors of single-authored documents, n (%)	38 (0.6)
International coauthorship (%)	26.88
Coauthors per document (average), n	5.57
Author's keywords, n	2917
References, n	45,050
Document age (average), n	5.7
Citations per document (average), n	21.95

Annual Scientific Production and Average Citations Per Year

Figure 2 shows that studies on exergaming began at very low levels in 1997, experienced significant growth starting in 2015, and peaked in 2021. Although there was a slight decline in 2022, scientific output continued in 2023 and 2024. It has been observed that the studies conducted in the field of exergaming and the average citation counts to these studies are not parallel, and there is a decrease in the average citation counts, unlike the number of publications.

Figure 2. Evaluation of exergaming studies by years and average citations per year.



XSL•FO
RenderX

The Most Effective Journals in the Field of Exergaming and Bradford's Law

An examination of the most relevant journals in the field of exergaming reveals that *Games for Health Journal* ranks first with 139 papers (Tables 2 and 3). This journal is followed by *JMIR Serious Games* with 63 papers. When evaluating the journals with the highest number of citations, *Games for Health Journal* once again tops the list with 1885 citations, followed by *Medicine & Science in Sports & Exercise* with 1241 citations. The local impact and impact factors of the top journals on exergaming are presented in Multimedia Appendix 3. According to the findings of Bradford's law, *Games for Health Journal* occupies a larger share among the core sources compared to other journals and is positioned within the first core group of journals. It is followed by *JMIR Serious Games* and the *International Journal of Environmental Research and Public Health*. The findings for Bradford's law are presented in Multimedia Appendix 4.

Table . Most relevant journals.

Most relevant journals	Papers, n
Games for Health Journal	139
JMIR Serious Games	63
International Journal of Environmental Research and Public Health	54
PLOS One	32
Journal of Clinical Medicine	28
Journal of NeuroEngineering and Rehabilitation	23
Frontiers in Psychology	22
Journal of Medical Internet Research	22
Frontiers in Aging Neuroscience	21
Applied Sciences—Basel	19

Table . Most cited journals.

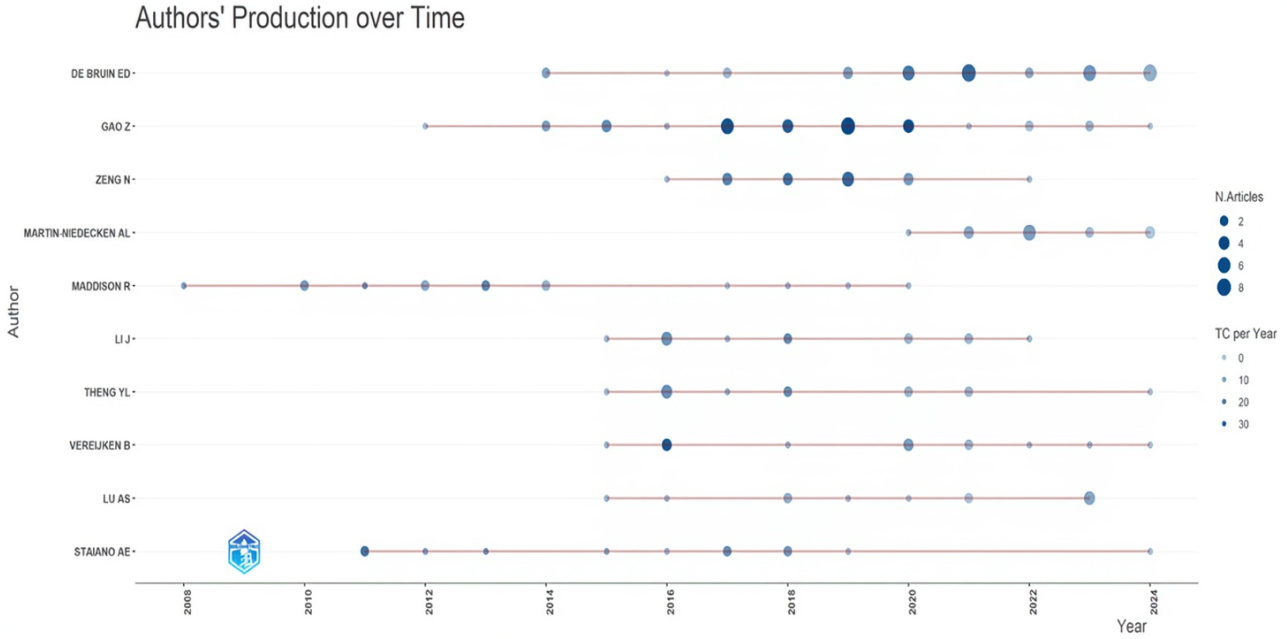
Most cited journals	Citation counts, n
Games for Health Journal	1885
Medicine and Science in Sports and Exercise	1241
PLOS One	929
Archives of Physical Medicine and Rehabilitation	882
Journal of NeuroEngineering and Rehabilitation	702
International Journal of Environmental Research and Public Health	693
Physical Therapy & Rehabilitation Journal	600
Computers in Human Behavior	558
Journal of Physical Activity and Health	548
Journal of Sport and Health Science	541

Authors' Productivity and the Distribution of Their Productivity Over the Years

Figure 3 shows the publication performance of the most productive authors on exergaming over the years. Gao Zan (n=36) and Eling de Bruin (n=36) are the most productive authors in this field and continue to work on related studies. The analysis of author productivity based on Lotka's law reveals that 79.74% (n=5226) of exergaming researchers have contributed with only 1 publication, 12.85% (n=842) have contributed with 2 publications, and 3.74% (n=245) with 3 publications. These findings do not fully align with Lotka's law. The findings on Lotka's law are presented in Multimedia Appendix 5.

XSL•FO
RenderX

Figure 3. Productivity of authors working on exergaming by year. TC: total citation.



Most Cited Sources

Table 4 shows the most cited papers on exergaming, demonstrating the impact and importance of research in this area. The most cited paper (TC=386) by Biddiss and Irwin [35]. The second most cited paper (TC=345) by Ijaz et al [5].

XSL•FO
RenderX

Table . Globally most cited documents on exergaming.

Title	Author and source	TC[a]	TCPY[b]	NTC[c]
1. Active video games to promote physical activity in children and youth: a systematic review	Biddiss and Irwin [35]	386	24.13	4.41
2. Player Experience of Needs Satisfaction (PENS) in an immersive virtual reality exercise platform describes motivation and enjoyment	Ijaz et al [5]	345	57.50	15.08
3. Exergaming and older adult cognition: a cluster randomized clinical trial	Anderson-Hanley et al [36]	304	21.71	5.18
4. Option exercise games: an application to the equilibrium investment strategies of firms	Grenadier [37]	298	12.42	1.00
5. Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games	Peng et al [10]	295	19.67	3.88
6. Effects of interactive physical-activity videogame training on physical and cognitive function in older adults	Maillot et al [38]	289	20.64	4.92
7. Exergames for physical education courses: physical, social, and cognitive benefits	Staiano and Calvert [39]	273	18.20	3.59
8. Need satisfaction supportive game features as motivational determinants: an experimental study of a self-determination theory guided exergame	Peng et al [40]	250	17.86	4.26
9. Playing active video games increases energy expenditure in children	Graf et al [41]	250	14.71	2.63
10. Efficacy and safety of non-immersive virtual reality exercising in stroke rehabilitation (EVREST): a randomised, multicentre, single-blind, controlled trial	Saposnik et al [42]	248	24.80	7.21

[a]TC: total citation.

[b]TCPY: total citation per year.

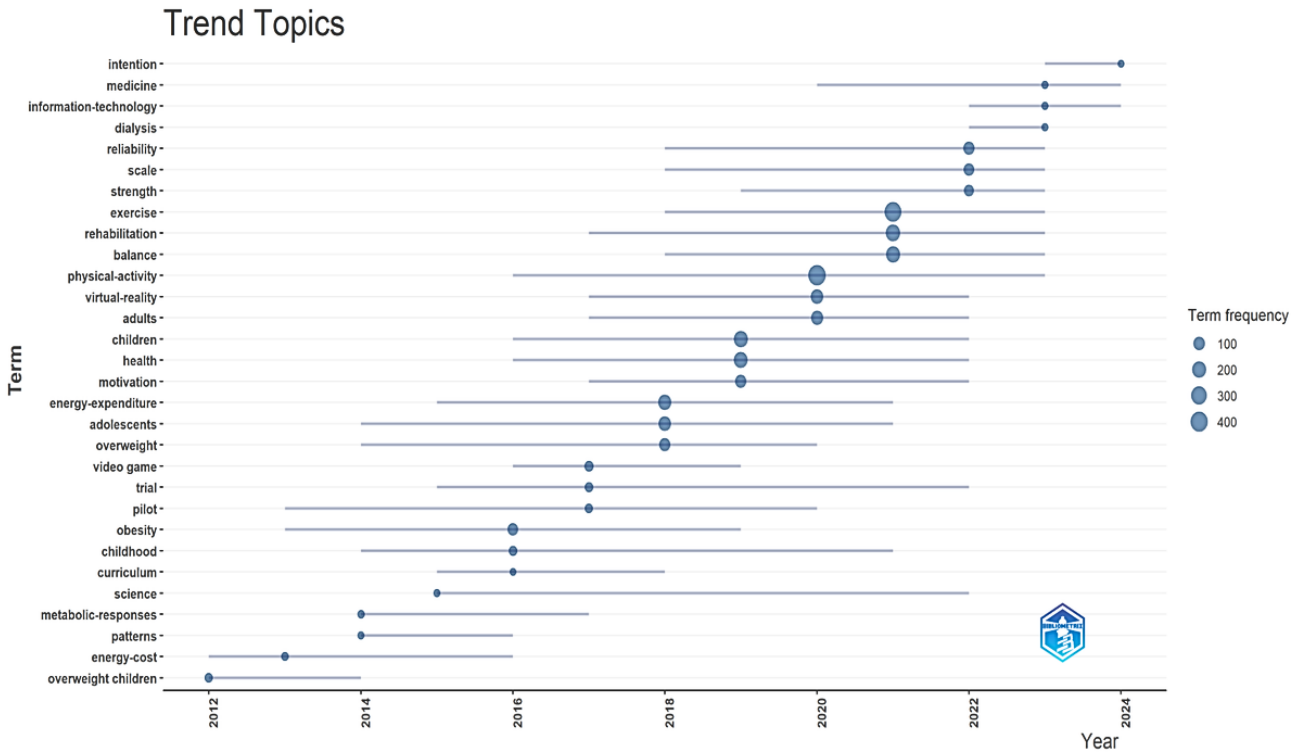
[c]NTC: normalized total citation.

Trending Topics

Figure 4 presents the trending topics in exergaming research over time, showing how these topics have changed over the years. Topics such as “health,” “physical activity,” “exercise,” “virtual reality,” and “obesity” have typically been focal points in this field. In recent years, topics such as “medicine,” “information technology,” and “intention” have emerged.

XSL•FO
RenderX

Figure 4. Popular exergaming topics by year.



Thematic Mapping

Figure 5 illustrates the development and centrality degrees of exergaming research themes. Motor themes represent well-developed and highly relevant core topics in exergaming research. No specific terms were identified in this region.

Basic themes encompass topics with high centrality but low density. These themes are foundational to research and represent broadly discussed areas. Notable terms in this region include “randomized controlled trial,” “virtual reality,” “physical activity,” “active video,” and “pilot study.”

Niche themes refer to topics with low centrality but high density. These themes represent specific research topics and specialized subfields within the domain. Terms such as “autism spectrum,” “randomized crossover,” and “neurocognitive disorder” are prominent in this region.

Emerging or declining themes are characterized by low centrality and low density. These themes represent either emerging or diminishing areas of interest in exergaming research. Notable terms in this region include “Wii fit” and “physical education.”

XSL•FO
RenderX

Figure 5. Thematic map findings for exergaming research.

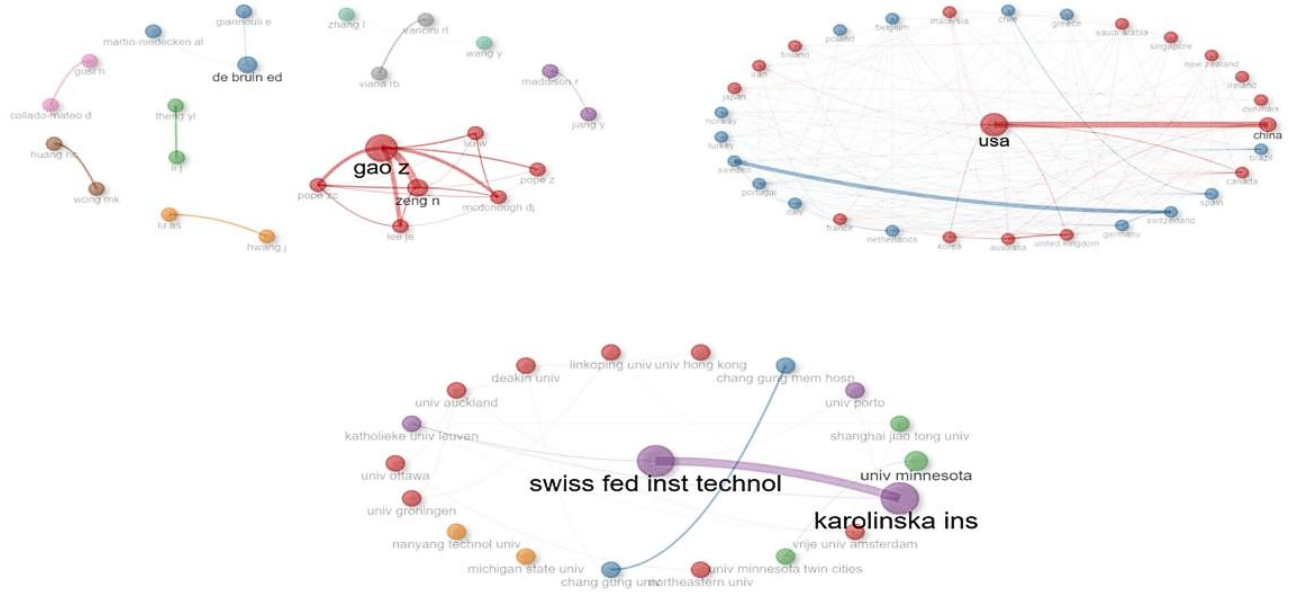


Collaboration

Figure 6 shows the most productive countries, the most productive authors, the most productive institutions, and the connections between them in exergaming research. One of the largest nodes on the graph, Gao Zan, is seen to collaborate closely with Zeng Nan, indicating that they have conducted many studies together. In terms of collaborations between countries, the United States is at the center, showing a strong collaboration with China. Leading institutions include the Swiss Federal Institute of Technology (ETH Zurich) and the Karolinska Institute, which have been very active and show strong relationships between them.

XSL•FO
RenderX

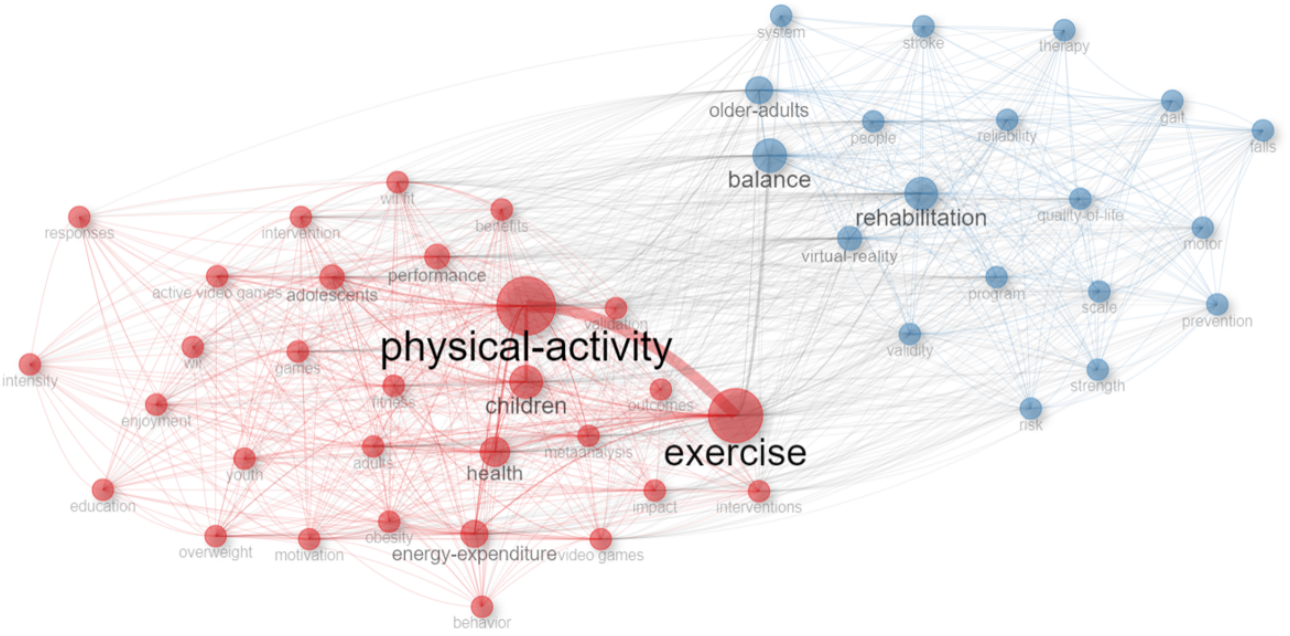
Figure 6. Collaborative network of researchers, countries, and institutions on exergaming.



Coword Network

Figure 7 presents a coword analysis based on the frequency of terms used in the titles, keywords, and abstracts of the 1626 included studies, highlighting the most prominent concepts in the field and the relationships between them. Using word cloud analysis in Bibliometrix, the main themes and concepts driving exergaming studies were identified, with PA (n=436), exercise (n=383), rehabilitation (n=218), and balance (n=213) ranking as the most prominent. The high frequency of these terms indicates their central role in exergaming research. Then coword analysis showed that the studies in the field are divided into 2 main clusters. Cluster 1 focuses on PA, exercise, performance, health, and health-related concepts, with these studies primarily targeting children and adolescents. Cluster 2, on the other hand, provides insights into studies focused predominantly on adults and older people, emphasizing rehabilitation, balance, and strength, highlighting the interconnectedness of these concepts.

Figure 7. Coword analysis of exergaming studies.



Discussion

Principal Findings

This bibliometric analysis has revealed a significant increase in exergaming research since 2010, peaking during the 2020‐2021 period. Thematic analyses demonstrate that exergaming research is inherently interdisciplinary and spans areas such as health, rehabilitation, and behavior change. Collaboration patterns and trend analyses highlight the potential of exergaming to address broader health challenges, providing critical insights to guide future research directions.

Annual Scientific Production in Exergaming Research

The annual production of scientific publications on exergaming has shown significant growth since 1997, following the release

XSL•FO
RenderX

of the Dance Dance Revolution. A prior bibliometric study focusing on VR, AR, and PA observed a similar growth trend, consistent with our findings [26,43]. A significant increase in publication volume has been observed since 2015, with scientific production peaking in 2020 - 2021. This growth reflects advancements in VR and AR technologies as well as increased investments in health-focused gaming applications and the growing role of technology-assisted interventions in health care services [21,44]. Furthermore, this rise can also be attributed to the COVID-19 pandemic, which highlighted the need for home-based solutions to maintain PA, thereby amplifying the significance of exergaming research [45,46]. After 2002, the average citation rate dropped sharply and did not follow a trajectory parallel to the increase in the number of publications. A possible explanation for this is that the growth in the number of studies within the exergaming field led to citations being distributed across a broader range of research. As the field expanded, citations may have been spread among a larger body of work, diminishing the citation impact of earlier studies.

Journals, Authors, and Papers Featured in Exergaming Research

Games for Health Journal, *International Journal of Environmental Research and Public Health*, and *JMIR Serious Games* are leading journals for exergaming research in health and education. These leading journals are essential platforms that facilitate the dissemination of innovative studies, promote interdisciplinary collaboration, and shape the applications of exergaming in improving health and PA outcomes [12,47,48]. According to Bradford's law, this finding indicates that most of the exergaming research is concentrated in a few core journals, while the remaining publications are distributed across a broader range of sources [49].

When examining the publication performance of the most prolific authors in the field of exergaming, a few notable names stand out, such as Gao Zan and Eling de Bruin. The continued production of relevant research by these authors, along with their numerous publications and high citation frequency, highlights their significant roles in the field of exergaming. According to Lotka's law, about 60% of researchers are expected to contribute with a single publication, 15% with 2 publications, and 6.67% with 3 publications [50]. Considering Lotka's law, which reveals the scientific productivity of authors in the field of exergaming, it can be concluded that researchers generally prefer to conduct a single study on this topic, and the number of authors specializing in the field remains limited. This shows that the literature in this field still has room for improvement.

The most cited on exergaming is Biddiss and Irwin's paper [35], which received 386 citations in total. This paper reviews the potential of exergaming to promote PA among children and adolescents and possibly lays the groundwork for further studies by emphasizing the positive effects of exergaming on child health [35]. Another influential paper is the paper by Ijaz et al [5], which received 345 citations and had the highest annual citation rate of 57.5. This study investigates how VR exergaming increases participants' motivation and enjoyment [5]. The high citation counts of these papers demonstrate their important contributions to the field of exergaming and their role as key reference points for other researchers.

Hot Spots in Exergaming Research

In the analysis of trending topics, early studies primarily concentrated on terms such as "obesity," "childhood," and "video games," whereas more specific topics like "virtual reality," "physical activity," and "rehabilitation" have gained prominence over time. This shift reflects the integration of advanced technologies and broader health applications into the field [51]. Furthermore, the rising frequency of terms such as "adolescents" and "adults" indicates a widening demographic focus, extending beyond children to include populations of different age groups. This evolution demonstrates the expanding scope of exergaming research, moving from a niche focus on childhood health to a broader application in preventive health care [52], rehabilitation [15], and lifestyle improvement. Finally, terms such as "intention" and "medicine" have more recently started to receive increasing attention, indicating that exergaming is being increasingly explored for behavior change and clinical interventions. Future studies could further enhance its impact by addressing less-explored topics and demographic groups.

Thematic analysis results reveal that fundamental themes, such as "virtual reality," "randomized controlled trials," and "physical activity," constitute the main research topics in exergaming, particularly focusing on health and rehabilitation applications.

Niche themes, including "crossover trials" and "autism spectrum disorders," delve into specialized research areas, allowing for in-depth studies on specific disease groups and research methods [53]. Themes such as "Wii Fit" and "physical education" efforts for school children represent topics that once gained popularity in exergaming research but have lost attention due to technological advancements and shifting research priorities. Future studies could explore niche themes in greater depth to discover new application areas and diversify basic themes to expand the potential of exergaming to broader audiences.

Author, Country, and Institutional Collaborations in Exergaming Research

Authors such as Gao Zan and Zeng Nan are notable for their extensive collaborations, which may stem from their shared interest in studying the effects of exergaming on health and PA. Furthermore, 26.88% (n=437.07) of the 1626 studies in the field of exergaming had international author collaborations, indicating that more than 1 researcher worked together.

The United States stands out as the most productive country, maintaining strong collaborations with many other nations. China also plays a crucial role, collaborating closely with the United States. These collaborations have facilitated extensive research on exergaming, supported by substantial scientific research funds, robust research infrastructures, and leading technological advancements. Our findings are consistent with bibliometric studies focusing on technologies like VR and AR, commonly used in exergaming applications [26,27].

Leading institutions like the Swiss Federal Institute of Technology (ETH Zurich) and Karolinska Institute are at the

XSL•FO
RenderX

forefront of this field. ETH Zurich, recognized for its expertise in technology and engineering, contributes considerably to the development and application of exergaming devices [54], while Karolinska Institute is prominent in medical research and clinical applications [55]. These institutional collaborations result in extensive and impactful studies investigating various applications of exergaming in fields such as health care and rehabilitation.

Future Research Suggestions

First, previous literature has shown that group-based exergames enhance motivation, promote sustained play, and improve self-efficacy. Family-friendly games like Ring Fit Adventure have the potential to increase PA while strengthening parent-child bonds. Future research should investigate the impact of these games on family dynamics and address critical issues such as "adherence," "sustainability," and "transfer effects." Second, significant increases in energy expenditure as a result of exergaming might be of little importance if one compensates by increasing energy intake. Therefore, promoting exergaming as an obesity prevention tool also requires the quantification of energy intake to adequately assess overall energy balance. Without considering both sides of the equation, energy expenditure and intake, it is difficult to determine the true effectiveness of exergaming in managing or preventing obesity. Future studies should incorporate both metrics to provide a clearer understanding of exergaming's potential health benefits. Third, exergaming has not been widely used in young children such as preschoolers; thus, it needs more research in this population. Fourth, with the advancement of technology and the introduction of new active video games, it is important to explore how these new games can promote PA and other health-related outcomes. Fifth, research could examine how exergaming can be integrated with technological advances and the potential benefits of this integration in health and education. Sixth, more research is needed on the social impacts and community-based applications of exergaming. Finally, further research on the economic issues and financial sustainability of exergaming is critical to assess its cost-effectiveness, economic benefits, and investment opportunities.

Limitations

This study has various limitations. First, only the WoS database was used because the robust analytical capabilities of the Bibliometrix R package with WoS data enhanced the reliability of our results, and WoS's "Keyword Plus" feature identified contextually significant keywords, thereby broadening the scope of the study. However, relying on a single database may exclude relevant sources from other databases, which is considered a limitation. Second, the R-based Bibliometrix tool was used for the analysis and visualization processes in this study, as it provided a sufficient platform aligned with the study's scope without incurring additional software costs. However, the exclusion of additional features offered by alternative software is considered a limitation of this study. Third, the metrics and indicators used, such as citation counts, journal impact factors, and authors' h-indexes, may not fully capture the quality of the research and potentially overestimate the impact of particular studies [56]. Finally, bibliometric methods provide an overview of a research field; however, this study does not offer an in-depth analysis of influential papers and authors. To gain a more comprehensive understanding of research impact and trends, bibliometric analyses should be complemented with other evaluation methods.

Conclusions

This study highlights the growing importance of exergaming research as a key indicator of the increasing integration of technology into the health domain. The COVID-19 pandemic has especially highlighted the value and relevance of home-based solutions in this field. The interdisciplinary nature of exergaming offers significant potential for the development of innovative approaches by bringing together technology and health applications. The thematic evolution of research demonstrates that exergaming has progressed beyond being merely a tool for children's entertainment or PA promotion, expanding into broader health contexts such as rehabilitation, behavior modification, and applications for diverse age groups. This trend suggests that exergaming could play a more active role in public health policies in the future. Future research focusing on underexplored demographic groups, emerging technologies, and niche health applications is likely to further enhance the impact of exergaming.

Acknowledgments

For some parts of the paper, we used ChatGPT (OpenAI) to improve the wording or to structure and enhance selected parts of the text. The final content has been carefully reviewed and approved by all authors, who take full responsibility for the accuracy and completeness of the work.

Data Availability

All data generated or analyzed during this study are included in this published paper.

Authors' Contributions

AK and EK contributed to the study conception and design. Material preparation, data collection, and analysis were performed by EK and AK. The first draft of the manuscript was written by AK, BK, and EK, and NZ revised and edited the manuscript. All authors read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy protocol.
[DOCX File, 13 KB - i-jmr_v14i1e66738_app1.docx]

Multimedia Appendix 2
Data from all included studies.
[XLS File, 7572 KB - i-jmr_v14i1e66738_app2.xls]

Multimedia Appendix 3
Local impact of the top journals on exergaming.
[PDF File, 70 KB - i-jmr_v14i1e66738_app3.pdf]

Multimedia Appendix 4
Scientific production of journals according to Bradford's law.
[PNG File, 103 KB - i-jmr_v14i1e66738_app4.png]

Multimedia Appendix 5
Author productivity findings according to Lotka's law.
[PDF File, 95 KB - i-jmr_v14i1e66738_app5.pdf]

References

1. Lee IM, Shiroma EJ, Lobelo F, et al. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229. [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
2. Bentlage E, Ammar A, How D, et al. Practical recommendations for maintaining active lifestyle during the COVID-19 Pandemic: a systematic literature review. Int J Environ Res Public Health 2020 Aug 28;17(17):32872154. [doi: 10.3390/ijerph17176265] [Medline: 32872154]
3. Oh Y, Yang S. Defining exergames & exergaming. 2010 Presented at: Proceedings of Meaningful Play; Oct 21-23, 2010; East Lansing, MI, United States.
4. Borodulin K, Sipilä N, Rahkonen O, et al. Socio-demographic and behavioral variation in barriers to leisure-time physical activity. Scand J Public Health 2016 Feb;44(1):62-69. [doi: 10.1177/1403494815604080] [Medline: 26392420]
5. Ijaz K, Ahmadpour N, Wang Y, Calvo RA. Player Experience of Needs Satisfaction (PENS) in an immersive virtual reality exercise platform describes motivation and enjoyment. Int J Hum-Comput Interact 2020 Aug 8;36(13):1195-1204. [doi: 10.1080/10447318.2020.1726107]
6. Ramírez-Granizo IA, Ubago-Jiménez JL, González-Valero G, Puertas-Molero P, San Román-Mata S. The effect of physical activity and the use of active video games: exergames in children and adolescents: a systematic review. Int J Environ Res Public Health 2020 Jun 14;17(12):32545877. [doi: 10.3390/ijerph17124243] [Medline: 32545877]
7. Zeng N, Pope Z, Gao Z. Acute effect of virtual reality exercise bike games on college students' physiological and psychological outcomes. Cyberpsychol Behav Soc Netw 2017 Jul;20(7):453-457. [doi: 10.1089/cyber.2017.0042] [Medline: 28715263]
8. Klos L, Feil K, Eberhardt T, Jekauc D. Interventions to promote positive affect and physical activity in children, adolescents and young adults—a systematic review. Sports (Basel) 2020 Feb 20;8(2):32093347. [doi: 10.3390/sports8020026] [Medline: 32093347]
9. García-Bravo S, Cuesta-Gómez A, Campuzano-Ruiz R, et al. Virtual reality and video games in cardiac rehabilitation programs. A systematic review. Disabil Rehabil 2021 Feb;43(4):448-457. [doi: 10.1080/09638288.2019.1631892] [Medline: 31258015]
10. Peng W, Lin JH, Crouse J. Is playing exergames really exercising? A meta-analysis of energy expenditure in active video games. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):681-688. [doi: 10.1089/cyber.2010.0578] [Medline: 21668370]
11. Ho RST, Chan EKY, Liu KKY, Wong SHS. Active video game on children and adolescents' physical activity and weight management: a network meta-analysis. Scand J Med Sci Sports 2022 Aug;32(8):1268-1286. [doi: 10.1111/sms.14176] [Medline: 35485251]
12. 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 Jan 14;10(1):e32331. [doi: 10.2196/32331] [Medline: 35029530]

XSL•FO
RenderX

13. Coknaz D, Mirzeoglu AD, Atasoy HI, Alkoy S, Coknaz H, Goral K. A digital movement in the world of inactive children: favourable outcomes of playing active video games in a pilot randomized trial. Eur J Pediatr 2019 Oct;178(10):1567-1576. [doi: 10.1007/s00431-019-03457-x] [Medline: 31471690]
14. Benzing V, Schmidt M. Exergaming for children and adolescents: strengths, weaknesses, opportunities and threats. J Clin Med 2018 Nov 8;7(11):30413016. [doi: 10.3390/jcm7110422] [Medline: 30413016]
15. Rüth M, Schmelzer M, Burtniak K, Kaspar K. Commercial exergames for rehabilitation of physical health and quality of life: a systematic review of randomized controlled trials with adults in unsupervised home environments. Front Psychol 2023;14(1155569):1155569. [doi: 10.3389/fpsyg.2023.1155569] [Medline: 37333591]
16. Peyrachon R, Rébillard A. Effects of active video games in patients with cancer: systematic review. JMIR Cancer 2023 May 26;9:e45037. [doi: 10.2196/45037] [Medline: 37234028]
17. Yun SJ, Hyun SE, Oh BM, Seo HG. Fully immersive virtual reality exergames with dual-task components for patients with Parkinson's disease: a feasibility study. J Neuroeng Rehabil 2023 Jul 18;20(1):92. [doi: 10.1186/s12984-023-01215-7] [Medline: 37464349]
18. Masoud AE, Shaheen AAM, Algabbani MF, AlEisa E, AlKofide A. Effectiveness of exergaming in reducing cancer-related fatigue among children with acute lymphoblastic leukemia: a randomized controlled trial. Ann Med 2023 Dec;55(1):2224048. [doi: 10.1080/07853890.2023.2224048] [Medline: 37318119]
19. Yang Y, Wang K, Liu S, Liu H, Zhang T, Luo J. Exergames improve cognitive function in older adults and their possible mechanisms: a systematic review. J Glob Health 2023 Dec 1;13:04177. [doi: 10.7189/jogh.13.04177] [Medline: 38038376]
20. Hastürk G, Akyıldız Munusturlar M. The effects of exergames on physical and psychological health in young adults. Games Health J 2022 Sep 20;11(6):36126314. [doi: 10.1089/g4h.2022.0093] [Medline: 36126314]
21. Donath L, Rössler R, Faude O. Effects of virtual reality training (exergaming) compared to alternative exercise training and passive control on standing balance and functional mobility in healthy community-dwelling seniors: a meta-analytical review. Sports Med 2016 Sep;46(9):1293-1309. [doi: 10.1007/s40279-016-0485-1] [Medline: 26886474]
22. Li J, Theng YL, Foo S. Effect of exergames on depression: a systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2016 Jan;19(1):34-42. [doi: 10.1089/cyber.2015.0366] [Medline: 26716638]
23. Zhang X, Chen H, Wang W, Ordóñez de Pablos P. What is the role of IT in innovation? A bibliometric analysis of research development in IT innovation. Behav Inform Technol 2016 Dec;35(12):1130-1143. [doi: 10.1080/0144929X.2016.1212403]
24. Dahesh MB, Tabarsa G, Zandieh M, Hamidizadeh M. Reviewing the intellectual structure and evolution of the innovation systems approach: a social network analysis. Technol Soc 2020 Nov;63:101399. [doi: 10.1016/j.techsoc.2020.101399]
25. Chuang KY, Huang YL, Ho YS. A bibliometric and citation analysis of stroke-related research in Taiwan. Scientometrics 2007 Aug;72(2):201-212. [doi: 10.1007/s11192-007-1721-0]
26. Denche-Zamorano A, Rodriguez-Redondo Y, Barrios-Fernandez S, et al. Rehabilitation is the main topic in virtual and augmented reality and physical activity research: a bibliometric analysis. Sensors (Basel) 2023 Mar 9;23(6):36991699. [doi: 10.3390/s23062987] [Medline: 36991699]
27. Pawassar CM, Tiberius V. Virtual reality in health care: bibliometric analysis. JMIR Serious Games 2021 Dec 1;9(4):e32721. [doi: 10.2196/32721] [Medline: 34855606]
28. Aria M, Cuccurullo C. bibliometrix: an R-tool for comprehensive science mapping analysis. J Informetr 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]
29. Boyack KW, Klavans R. Co‐citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately? J Am Soc Inf Sci 2010 Dec;61(12):2389-2404. [doi: 10.1002/asi.21419]
30. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res 2021 Sep;133:285-296. [doi: 10.1016/j.jbusres.2021.04.070]
31. Su X, Li X, Kang Y. A bibliometric analysis of research on intangible cultural heritage using CiteSpace. SAGE Open 2019 Apr;9(2):1-18. [doi: 10.1177/2158244019840119]
32. Paez A. Gray literature: an important resource in systematic reviews. J Evid Based Med 2017 Aug;10(3):233-240. [doi: 10.1111/jebm.12266] [Medline: 28857505]
33. 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]
34. Jiang Y, Ritchie BW, Benckendorff P. Bibliometric visualisation: an application in tourism crisis and disaster management research. Curr Issues Tour 2019 Oct 2;22(16):1925-1957. [doi: 10.1080/13683500.2017.1408574]
35. Biddiss E, Irwin J. Active video games to promote physical activity in children and youth: a systematic review. Arch Pediatr Adolesc Med 2010 Jul;164(7):664-672. [doi: 10.1001/archpediatrics.2010.104] [Medline: 20603468]
36. Anderson-Hanley C, Arciero PJ, Brickman AM, et al. Exergaming and older adult cognition: a cluster randomized clinical trial. Am J Prev Med 2012 Feb;42(2):109-119. [doi: 10.1016/j.amepre.2011.10.016] [Medline: 22261206]
37. Grenadier SR. Option exercise games: an application to the equilibrium investment strategies of firms. Rev Financ Stud 2002 Apr;15(3):691-721. [doi: 10.1093/rfs/15.3.691]
38. Maillot P, Perrot A, Hartley A. Effects of interactive physical-activity video-game training on physical and cognitive function in older adults. Psychol Aging 2012 Sep;27(3):589-600. [doi: 10.1037/a0026268] [Medline: 22122605]

XSL•FO
RenderX

39. Staiano AE, Calvert SL. Exergames for physical education courses: physical, social, and cognitive benefits. Child Dev Perspect 2011 Jun;5(2):93-98. [doi: 10.1111/j.1750-8606.2011.00162.x] [Medline: 22563349]
40. Peng W, Lin JH, Pfeiffer KA, Winn B. Need satisfaction supportive game features as motivational determinants: an experimental study of a self-determination theory guided exergame. Media Psychol 2012 May 18;15(2):175-196. [doi: 10.1080/15213269.2012.673850]
41. Graf DL, Pratt LV, Hester CN, Short KR. Playing active video games increases energy expenditure in children. Pediatrics 2009 Aug;124(2):534-540. [doi: 10.1542/peds.2008-2851] [Medline: 19596737]
42. Saposnik G, Cohen LG, Mamdani M, et al. Efficacy and safety of non-immersive virtual reality exercising in stroke rehabilitation (EVREST): a randomised, multicentre, single-blind, controlled trial. Lancet Neurol 2016 Sep;15(10):1019-1027. [doi: 10.1016/S1474-4422(16)30121-1] [Medline: 27365261]
43. Fan T, Wang X, Song X, Zhao G, Zhang Z. Research status and emerging trends in virtual reality rehabilitation: bibliometric and knowledge graph study. JMIR Serious Games 2023 Mar 6;11:e41091. [doi: 10.2196/41091] [Medline: 36877556]
44. Ng YL, Ma F, Ho FK, Ip P, Fu KW. Effectiveness of virtual and augmented reality-enhanced exercise on physical activity, psychological outcomes, and physical performance: a systematic review and meta-analysis of randomized controlled trials. Comput Human Behav 2019 Oct;99:278-291. [doi: 10.1016/j.chb.2019.05.026]
45. Ellis LA, Lee MD, Ijaz K, Smith J, Braithwaite J, Yin K. COVID-19 as "game changer" for the physical activity and mental well-being of augmented reality game players during the pandemic: mixed methods survey study. J Med Internet Res 2020 Dec 22;22(12):e25117. [doi: 10.2196/25117] [Medline: 33284781]
46. Campo-Prieto P, Rodríguez-Fuentes G, Cancela-Carral JM. Immersive virtual reality exergame promotes the practice of physical activity in older people: an opportunity during COVID-19. Multimodal Technol Interact 2021;5(9):52. [doi: 10.3390/mti5090052]
47. Qian JL, McDonough DJ, Gao Z. The effectiveness of virtual reality exercise on individual's physiological, psychological and rehabilitative outcomes: a systematic review. Int J Environ Res Public Health 2020 Jun 10;17(11):4133. [doi: 10.3390/ijerph17114133] [Medline: 32531906]
48. Özlü A, Üstündağ S, Bulut Özkaya D, Menekşeoğlu AK. Effect of exergame on pain, function, and quality of life in shoulder impingement syndrome: a prospective randomized controlled study. Games Health J 2024 Apr;13(2):109-119. [doi: 10.1089/g4h.2023.0108] [Medline: 38394299]
49. Cobo MJ, López-Herrera AG, Herrera-Viedma E, Herrera F. Science mapping software tools: review, analysis, and cooperative study among tools. J Am Soc Inf Sci 2011 Jul;62(7):1382-1402. [doi: 10.1002/asi.21525]
50. Lotka AJ. The frequency distribution of scientific productivity. J Wash Acad Sci 1926;16(12):317-323 [FREE Full text]
51. Xu WG, Liang HN, Yu KY, Wen SY, Baghaei N, Tu HW. Acceptance of virtual reality exergames among Chinese older adults. Int J Hum-Comput Interact 2023 Mar 16;39(5):1134-1148. [doi: 10.1080/10447318.2022.2098559]
52. Sadeghi H, Jehu DA. Exergaming to improve physical, psychological and cognitive health among home office workers: a COVID-19 pandemic commentary. Work 2022;71(1):13-17. [doi: 10.3233/WOR-211000] [Medline: 34924430]
53. da Silva Alves R, de Carvalho JM, Borges JBC, Nogueira DA, Iunes DH, Carvalho LC. Effect of exergaming on quality of life, fatigue, and strength and endurance muscle in cancer patients: a randomized crossover trial. Games Health J 2023 Oct 1;12(5):358-365. [doi: 10.1089/g4h.2022.0161]
54. Manser P, de Bruin ED. Making the best out of it: design and development of exergames for older adults with mild neurocognitive disorder—a methodological paper. Front Aging Neurosci 2021;13(734012):734012. [doi: 10.3389/fnagi.2021.734012] [Medline: 34955806]
55. Adcock M, Fankhauser M, Post J, et al. Effects of an in-home multicomponent exergame training on physical functions, cognition, and brain volume of older adults: a randomized controlled trial. Front Med 2020;6:321. [doi: 10.3389/fmed.2019.00321]
56. Romanelli JP, Gonçalves MCP, de Abreu Pestana LF, Soares JAH, Boschi RS, Andrade DF. Four challenges when conducting bibliometric reviews and how to deal with them. Environ Sci Pollut Res 2021 Nov;28(43):60448-60458. [doi: 10.1007/s11356-021-16420-x]

Abbreviations

AR: augmented reality
PA: physical activity
RQ: research question
TC: total citation
VR: virtual reality
WoS: Web of Science

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 21.09.24; peer-reviewed by J Gu, R Gorantla; revised version received 10.04.25; accepted 16.04.25; published 25.08.25.

Please cite as:
Kayhan A, Kose E, Kayhan B, Zeng N
Mapping Hot Spots and Global Research Trends in Exergaming Between 1997 and 2024: Bibliometric Analysis
Interact J Med Res 2025;14:e66738
URL: https://www.i-jmr.org/2025/1/e66738
doi:10.2196/66738

© Abdullah Kayhan, Elif Kose, Burcu Kayhan, Nan Zeng. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 25.8.2025. 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

Integration of Conventional and Virtual Reality Approaches in Augmented Reality for Theory-Based Psychoeducational Intervention Design for Chronic Low Back Pain: Scoping Review

Robin Conen[1], MSc; Steffen Mueller[2], Prof Dr; Ana Nanette Tibubos[1,3], Prof Dr

[1]Department of Nursing Science, Diagnostics in Healthcare and eHealth, Trier University, Trier, Germany

[2]Department of Computer Science/Therapeutic Science, Trier University of Applied Sciences, Trier, Germany

[3]Department of Psychosomatic Medicine and Psychotherapy, University Medical Center of the Johannes Gutenberg-University Mainz, Mainz, Germany

Corresponding Author:
Robin Conen, MSc
Department of Nursing Science, Diagnostics in Healthcare and eHealth
Trier University
Max-Planck-Straße 6
Trier, 54296
Germany
Phone: 49 651 201 1904
Email: conen@uni-trier.de

Abstract

Background: Psychoeducation positively influences the psychological components of chronic low back pain (CLBP) in conventional treatments. The digitalization of health care has led to the discussion of virtual reality (VR) interventions. However, CLBP treatments in VR have some limitations due to full immersion. In comparison, augmented reality (AR) supplements the real world with virtual elements involving one's own body sensory perception and can combine conventional and VR approaches.

Objective: The aim of this study was to review the state of research on the treatment of CLBP through psychoeducation, including immersive technologies, and to formulate suggestions for psychoeducation in AR for CLBP.

Methods: A scoping review following PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines was performed in August 2024 by using Livivo ZB MED, PubMed, Web of Science, American Psychological Association PsycINFO (PsycArticle), and PsyArXiv Preprints databases. A qualitative content analysis of the included studies was conducted based on 4 deductively extracted categories.

Results: We included 12 studies published between 2019 and 2024 referring to conventional and VR-based psychoeducation for CLBP treatment, but no study referred to AR. In these studies, educational programs were combined with physiotherapy, encompassing content on pain biology, psychological education, coping strategies, and relaxation techniques. The key outcomes were pain intensity, kinesiophobia, pain catastrophizing, degree of disability, quality of life, well-being, self-efficacy, depression, attrition rate, and user experience. Passive, active, and gamified strategies were used to promote intrinsic motivation from a psychological point of view. Regarding user experience from a software development perspective, user friendliness, operational support, and application challenges were recommended.

Conclusions: For the development of a framework for an AR-based psychoeducational intervention for CLBP, the combination of theories of acceptance and use of technologies with insights from health psychological behavior change theories appears to be of great importance. An example of a theory-based design of a psychoeducation intervention in AR for CLBP is proposed and discussed.

(Interact J Med Res 2025;14:e59611) doi:10.2196/59611

KEYWORDS

augmented reality; virtual reality; chronic low back pain; education; pain management; intervention

XSL•FO
RenderX

Introduction

Globally, 60%-80% of adults experience low back pain, with 10% developing chronic forms, of which 85% are classified as chronic nonspecific low back pain without a clear etiology [1]. Owing to the limited efficacy and adverse effects of pharmacological approaches, there is a need for nonpharmacological alternatives [2] to improve treatment outcomes [3] and develop effective behavioral interventions [4]. Treatment guidelines recommend behavioral modification, exercise, psychoeducation [5-7], and physiotherapy for trunk muscle strengthening [8-11] to reduce pain and disability.

Educational interventions for chronic low back pain (CLBP) provide knowledge about the condition, coping strategies, and physical activity [3,12], with the objective of enhancing the quality of life and symptom management by mitigating anxiety, kinesiophobia, hyperactive pain behavior, and depression, which are risk factors for pain chronification. Additionally, psychoeducation fosters self-efficacy to break the cycle between anxiety and pain [13,14].

Many traditional interventions to boost physical activity, which is key for CLBP treatment, rely on intention theories for modifying health behavior [15]. A prominent intention theory is the Unified Theory of Acceptance and Use of Technology 2 (UTAUT 2) by Venkatesh et al [16], which examines the acceptance and use of technologies. It has gained recognition in fields such as education, e-commerce, and health research with advancing health technology [17,18]. UTAUT 2 explains the formation of intention for technology use through the constructs of performance expectancy, effort expectancy, social influence, facilitating conditions, hedonic motivation, price value, and habit. These factors may also be useful for predicting the intentions of patients with CLBP toward educational technology adoption. Furthermore, insights from theories focusing on health behavior change may prove fruitful to consider when attempting to change health behavior through the use of a new technology. For instance, Schwarzer's Health Action Process Approach model [19], commonly used in health behavior interventions, highlights self-efficacy and outcome expectation. The Health Action Process Approach distinguishes between intention formation and implementation as well as between nonintenders, intenders, and actors, each requiring tailored interventions to promote self-efficacy, information, and support in implementation [20]. Another example is Michie's Behavior Change Technique (BCT) taxonomy with 93 BCTs outlining strategies for successful behavior change [21].

Immersive technologies can be characterized on the Reality-Virtuality Continuum by Milgram and Kishino [22]. They demonstrate visual display technologies ranging from real to virtual environments, including augmented reality (AR) and virtual reality (VR) [23-25]. AR enables the concurrent presence and interaction of digital and physical elements within real-world environments in real time. VR, in contrast, enables complete immersion in VR and represents the extreme of Milgram's continuum between reality and virtuality [26].

With regard to research in immersive technologies such as VR in the treatment of CLBP, VR-based treatments turned out to be promising in reducing acute, experimental, and chronic pain and can complement conventional CLBP treatments [27].

VR has proven effective in treating acute pain [24] by redirecting attention from unpleasant stimuli such as back pain to more pleasant visual, auditory, and tactile stimuli [27]. VR interventions were found to reduce pain intensity, catastrophizing symptoms, and psychological symptoms in patients with CLBP after one session through distraction, indicating the direct influence of VR on pain perception [26]. Other VR studies demonstrated the feasibility and efficacy of VR for CLBP as an alternative approach, such as VR applications with graded exposure during walking and grasping with integrated game design [28], self-administered VR therapy for CLBP at home [29], and its implementation even during COVID-19 [30]. A recent meta-analysis also showed that kinesiophobia and pain intensity in CLBP can be reduced through VR training [31].

Although there is some evidence for the safety and tolerability of VR treatment for CLBP, most studies lack methodological quality and results were limited to short-term effects. Studies on safety, acceptance, and satisfaction are lacking, including targeted investigations of the risks of spinal pain caused by VR [32]. Thus, while VR is promising in reducing CLBP symptoms, AR might offer additional benefits through the integration of physical and virtual elements, thereby reducing VR-associated discomfort. AR enables the coexistence and interaction of virtual and physical objects in real time in the real world, thus combining the advantages of VR while mitigating its limitations such as cybersickness and visual discomfort [33]. AR can enhance interaction, presence, intuitiveness, and pedagogical flexibility by enriching the real world with digital information, accommodating various learning styles, and facilitating teaching and learning [34]. Despite these presumed advantages of AR, to our knowledge, there are no empirical studies of AR-based treatment for CLBP.

In summary, pain treatment guidelines emphasize the key role of educational CLBP treatment to counteract psychological chronification and promote self-efficacy according to health behavior change models. Furthermore, when health behavior change is addressed using a new technology, a joint consideration of health psychological models with theories of acceptance from a technological perspective, like the UTAUT 2 [35], is considered useful for successful implementation. Existing studies with immersive technology [24,26-31] demonstrated positive effects for CLBP treatment in VR incorporating psychoeducational elements. However, these VR studies have methodological shortcomings and gaps regarding dimensions of user experience such as satisfaction and acceptance. Therefore, by formulating research questions using the PICO (Population, Intervention, Comparison, Outcome) framework, this scoping review aims to first examine research in patients with CLBP (P) receiving psychoeducation through immersive technology (I) compared to conventional psychoeducation (C) to improve pain relief and pain-psychological variables (O) and second on the basis of the results of the literature analysis, to develop an intervention design for AR-based psychoeducation in patients with CLBP that combines conventional methods with immersive technology

XSL•FO
RenderX

based on a technology acceptance model to promote acceptance and pain management.

Methods

We investigated the research question through a scoping review and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews [36]. This review includes studies that used psychoeducation for CLBP and chronic pain treatment: (1) conventionally, (2) with immersive technology in VR or AR, or (3) a combination of both, conventional therapy with VR or AR technology use. Only papers published in English or German in 2019-2024 were considered, wherein clinical guidelines were generally updated every 3-5 years with new evidence [37]. The exclusion criteria were as follows: (1) psychiatric patients, (2) acute back pain, (3) back pain after medical procedures, and (4) other specific pain conditions and pharmacological interventions. Scientific investigations or studies in journals or textbooks were included, regardless of the scientific methodology used. An electronic search was performed in August 2024 by using predefined English terms: ("chronic low back pain" OR "CLBP" OR "chronic pain") AND (("virtual reality" OR "augmented reality") OR ("education" OR "multimodal pain therapy" OR "psychological intervention")). Reviewer RC used Citavi to search for in vivo ZB MED and PubMed, and a manual search was conducted in the Web of Science, American Psychological Association PsycINFO, and PsyArXiv Preprints. The search was conducted in line with the Joanna Briggs Institute methodology for scoping reviews, extending the PRISMA statement [38]. In accordance with Arksey and O'Malley's [39] recommendations for scoping reviews, we did not include a formal quality assessment of the incorporated research. The selection process was initially based on a review of titles and abstracts regarding the inclusion and exclusion criteria, followed by an assessment of the full text by a reviewer (RC) and double-checked by another reviewer (ANT). Both reviewers (RC and ANT) extracted the following information from the included studies by using Microsoft Excel, following the Joanna Briggs Institute model: (1) citation, (2) context, (3) participant characteristics, (4) study aim, (5) methodology, (6) results, (7) interventions, (8) limitations, (9) key results related to review questions, and (10) future research areas [40]. Data analysis by the first reviewer (RC) utilized a qualitative content analysis [41]. A deductive approach was used to extract relevant categories for achieving the research objective. Guidelines for the treatment of CLBP [42] as well as recommendations of the World Health Organization for digital health interventions [43] served as the basis for this. Subsequently, 4 categories were extracted to capture all the essential aspects relevant to the design of the envisaged intervention. The categories are as follows: (1) content of CLBP-specific education, (2) factors based on the psychology of learning for the intervention design, (3) technical conditions (framework) for CLBP interventions, and (4) outcome measures of the educational interventions for CLBP.

Results

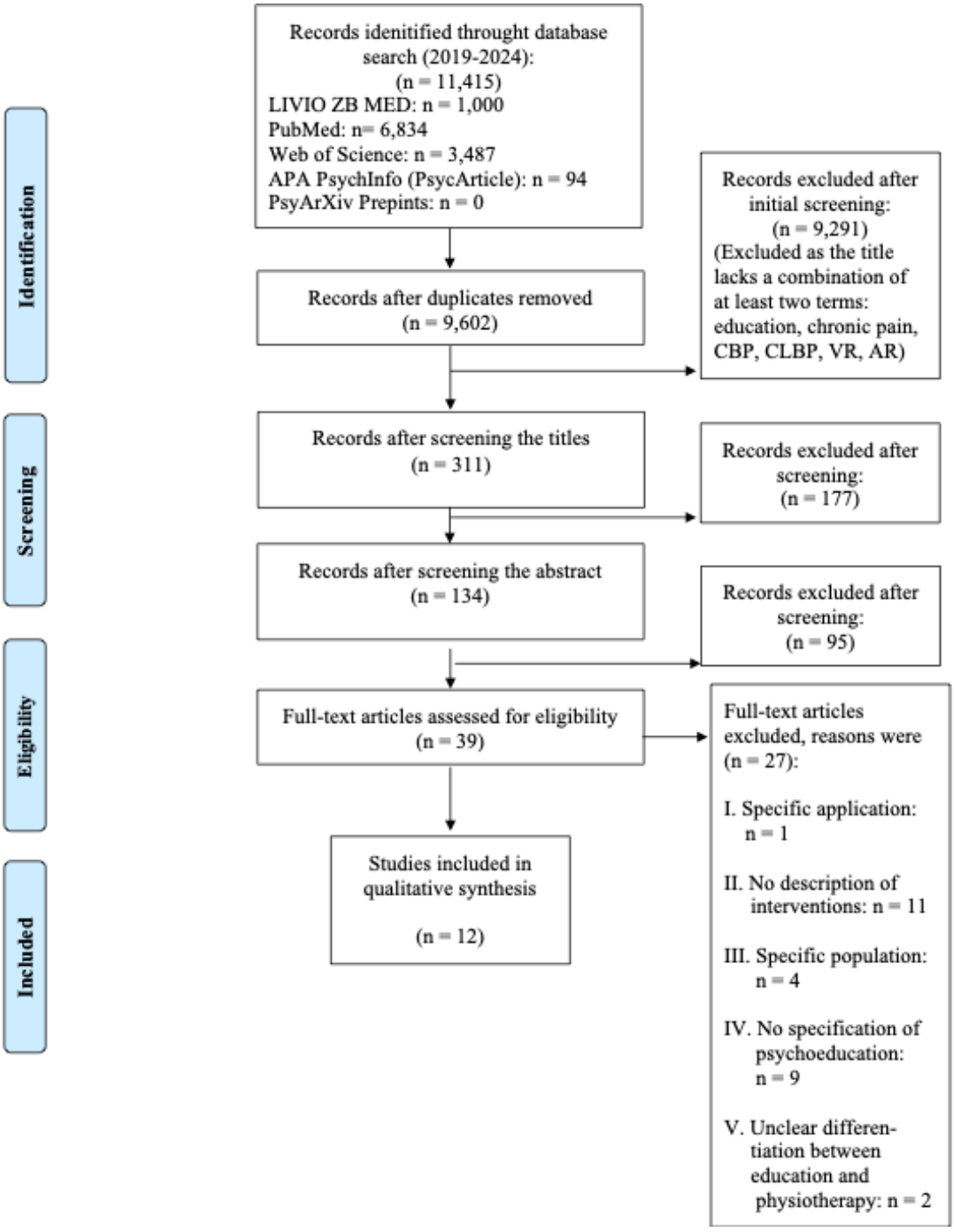
Study Selection

The study selection process, as shown in Figure 1, began with a database search that yielded 11,415 results. A total of 9602 titles were screened for the following terms: education, chronic pain, chronic back pain, CLBP, VR, and AR. The title should contain a minimum of 2 of the following keywords: education, chronic pain, chronic back pain, CLBP, VR, or AR; 9291 papers were excluded due to the lack of appearance of at least 2 of the defined terms. After applying the inclusion and exclusion criteria to 311 publications, 177 studies were excluded based on the titles and 95 were excluded based on the abstracts. After reviewing the full texts of the remaining 39 publications, 27 were excluded for (1) specific applications (eg, doctor-patient communication), (2) insufficient intervention descriptions, (3) overly specific populations (eg, elite athletes, primary school students, nursing staff), (4) unspecified psychoeducation (eg, cognitive behavioral therapy [CBT], cognitive functional training), and (5) unclear differentiation between education and physiotherapy in intervention design. Finally, the scoping review analyzed 12 publications, displayed in Table 1 [44-55].

The studies originated from Italy, Spain, the Netherlands, France, Chile, India, and Tunisia (8.33% each), Germany (25%), and the United States (16.67%). The review included 9 empirical studies (1 interview study, 8 interventional studies) and 3 reviews (1 systematic, 1 scoping, and 1 narrative review). All studies were peer-reviewed, except the narrative review. The results of the analysis of the 12 publications included are presented below according to the 4 defined categories.

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the study selection process. APA: American Psychological Association; AR: augmented reality; CBP: chronic back pain; CLBP: chronic low back pain; VR: virtual reality.



XSL•FO
RenderX

Table 1. Included studies applying psychoeducation by using conventional approaches and virtual reality approaches for the treatment of chronic low back pain [44-55].

Author	Study title	Journal name
Salazar-Méndez et al [44], 2024	Pain Neuroscience Education for Patients With Chronic Pain: A Scoping Review From Teaching-Leaning Strategies, Educational Level, and Cultural Perspective	Patient Education and Counseling
Ferlito et al [45], 2022	Pain Education in the Management of Patients with Chronic Low Back Pain: A Systematic Review	Journal of Functional Morphology and Kinesiology
Rim et al [46], 2022	Efficiency of Associating Therapeutic Patient Education with Rehabilitation in the Management of Chronic Low Back Pain: A Randomized Controlled Trial	Korean Journal of Family Medicine
Sidiq et al [47], 2024	Effects of Pain Education on Disability, Pain, Quality of Life, and Self-Efficacy in Chronic Low Back Pain: A Randomized Controlled Trial	PLOS One
Tomás-Rodríguez et al [48], 2024	Short- and Medium-Term Effects of a Single Session of Pain Neuroscience Education on Pain and Psychological Factors in Patients With Chronic Low Back Pain: A Single-Blind Randomized Clinical Trial	European Journal of Pain
Janik et al [49], 2024	Middle-Term Effects of Education Program in Chronic Low Back Pain Patients to an Adherence to Physical Activity: A Randomized Controlled Trial	Patient Education and Counseling
Lindner et al [50], 2020	Use of Virtual Reality as a Component of Acute and Chronic Pain Treatment	Anasthesiologie Intensivmedizin Notfallmedizin Schmerztherapie
Stamm et al [51], 2020	Virtual Reality in Pain Therapy: a Requirements Analysis for Older Adults With Chronic Back Pain	Journal of NeuroEngineering and Rehabilitation
Stamm et al [52], 2022	Virtual Reality Exergame for Supplementing Multimodal Pain Therapy in Older Adults With Chronic Back Pain	Virtual Reality
Brown et al [53], 2023	Chronic Pain Education Delivered With a Virtual Reality Headset in Outpatient Physical Therapy Clinics: A Multisite Exploratory Trial	American Journal of Translational Research
McConnell et al [54], 2024	A Multicenter Feasibility Randomized Controlled Trial Using a Virtual Reality Application of Pain Neuroscience Education for Adults With Chronic Low Back Pain	Annals of Medicine
de Vries et al [55], 2023	Pain Education and Pain Management Skills in Virtual Reality in the Treatment of Chronic Low Back Pain: A Multiple Baseline Single-Case Experimental Design	Behavior Research and Therapy

Contents of CLBP-Specific Education

The systematic review evaluated clinical studies from 2011 to 2021 comparing pain education/CBT with conventional physiotherapy for CLBP [45]. Thirteen studies, including 12 randomized controlled trials with 1642 participants, were analyzed. Six studies demonstrated a significant reduction in pain compared with the control group. The review concluded that due to the multimodality and heterogeneity of treatments, no definitive statement can be made regarding the efficacy of pain education or CBT in patients with CLBP [45]. Seven studies included an educational program in conjunction with physiotherapy [45-48,52-54]. Educational content varied considerably, ranging from exclusive focus on pain biology [46-49,52,53,55] to the inclusion of psychological aspects [46,52-55] and multidisciplinary approaches [49]. Most of the programs [48,55] incorporated education on pain physiology, frequently based on the book "Explain Pain" [56] by Butler and Moseley. Psychological education encompassed topics such as physical activity [45,46,49,53], fear of physical activity, emotional management [45-49,51], lifestyle modifications, daily exercises [45,49,52], pain-specific coping strategies [45,52,54], pain sensitization [47,54,55], and relaxation techniques, including stress management and mindfulness [45,49,52-55] in 6 studies, of which 5 included VR interventions. The content, duration, and physiotherapeutic integration of the individual education programs can be found in Multimedia Appendix 1.

Factors Based on Psychology of Learning for Intervention Design

This category encompasses factors of psychology of learning that are pertinent to the design of interactive interventions. Analysis of 3 studies showed that VR-based education employs passive mediation strategies such as informational videos and lectures (provided conventionally and in VR), alongside active and interactive strategies [44,50,54]. Three studies mentioned VR-based gamified approaches [50,52,55] and 2 studies [50,51] mentioned the promotion of intrinsic motivation.

Mediation Strategies

The included systematic review [44] examined the programs, cultural adaptations, and the efficacy of pain neuroscience education for chronic musculoskeletal pain, analyzing 71 studies that met our inclusion criteria and featured pain duration exceeding 3 months in adults. The analyzed studies explored pain neuroscience education in different settings by using

XSL•FO
RenderX

various experimental designs, including secondary analyses of randomized controlled trials, and showed positive effects on pain and psychological variables. Despite cultural influences on pain-relevant factors, only 2 (3%) of the 71 studies culturally adapted the pain neuroscience education material. Passive teaching-learning strategies tended to yield better outcomes for pain and functionality, whereas active methods resulted in significant knowledge improvements, albeit with insufficient description. The outcomes of multimodal therapies for chronic pain depend on the individualized integration of pain-specific education, considering biopsychosocial factors, educational level, culture, and diverse learning methods and materials for conveying pain neuroscience content [44]. Interaction content is passively conveyed through videos or lectures [46-49,53], particularly VR-based 360° nature videos [53].

Gamification and Motivation Enhancement

The Pain-Neuro-Education 2.0 software utilized a VR headset with immersive footage and computer-generated images for visually and emotionally engaging educational and relaxation training for chronic pain. This included interactive emotion regulation exercises such as breathing and mindfulness exercises in natural environments [54]. The VR program Recupt was also used to convey information in an engaging manner by having the user shoot at the pain stimulus with a laser gun, among other things. In the spinal cord phase, participants focus on visual “pain gates” and breathing to metaphorically “close” them and experience relaxation-induced pain relief. The brain component elucidates the reduction in pain response through the visualization and reactivation of illuminated connections. The alarm center gameplay demonstrates how emotions, cognitions, and behaviors influence pain perception. Finally, participants envision the alarm center as a brain region that regulates pain stimuli in an aircraft cockpit [55]. The VR program ViRST provides a therapeutic, interactive user interface with task-based activities in a farm environment [52]. Patients visualize movements and exertion levels by using game-based biofeedback with progress tracking and narrative elements [51]. Exergames incorporate biofeedback such as heart rate variability via photoplethysmography to prevent overexertion in interactive scenarios [52]. Gamification can motivate and enhance therapy adherence by fulfilling the psychological needs of competence, autonomy, and relatedness through interactive knowledge transfer. It also improves user skills through playful activities [50]. In the long term, feedback should be framed positively to maintain intrinsic motivation [52]. Avatars manipulate body perception for therapeutic effects, with the Proteus effect causing users to adopt their avatar’s behavior in real life. Personalized avatars can amplify pain relief [50].

Technical Conditions (Framework) for CLBP Intervention

The technical parameters of 3 enclosed VR studies, comprising 1 needs analysis [52] and 2 feasibility studies [52,53], provide insights into the design of AR-based education and identify potential areas of focus such as user-friendliness [51,53], operational support [51-53], and various application challenges [51,53]. The needs analysis was based on semistructured interviews (n=10) in focus groups to determine the requirements of older patients with chronic back pain, physiotherapists, and psychotherapists regarding VR pain therapy in terms of overall system, hardware, and software [51]. Findings emphasize that the designed system must be user-friendly; provide personalized instructions, demonstration videos, and individual guidance; and be available for rent. Assistants should support this system. Automatic breaks were considered crucial to avoid overexertion and pain aggravation. Activity should be limited to 30 minutes followed by a 15-minute rest. The study also highlighted the importance of balancing active therapy and relaxation. For hardware, it was determined that the VR headset must be independent and removable. Software design should consider user-friendliness by integrating the game environment with the level in-game environment for individual calibration of movement restrictions, particularly in gaming activities. Finally, a spacious room and wireless head-mounted display were considered essential for safety to prevent falls.

One feasibility study also emphasized the importance of safety aspects for the usability of VR headsets. The authors indicate that 93% of the application issues were associated with handling spatial and temporal limitations [53]. The second feasibility study demonstrated that disregarding body height (insufficient arm span) was perceived as disruptive [52]. Operational support software allows therapists to intervene during instances of pain, anxiety, or improper exercise execution by using a help button or emergency assistance [51]. Incorrect exercise execution is considered disruptive [52] and often lacks adequate support personnel for error correction or clinical assistance [53].

Outcome Measures of Educational Interventions for CLBP

Of the 12 studies [44-55] reviewed, 8 [46-49,54,55] were quantitative interventional studies. Commonly evaluated outcomes in CLBP studies encompassed pain intensity [45-49,51,52,54,55], kinesiophobia [46,48,51-53,55], pain catastrophizing [48,53-55], disability [45-47,51,52,54], health-related quality of life [51,52,54], well-being [47], self-efficacy [47,54], depression [46,53], attrition rate [49,53], and VR intervention user experience [52,53]. A comprehensive table displaying the different methods used, study results, and conclusions for each of the included studies can be found in Multimedia Appendix 2. A list of the pain-specific constructs and their measurement tools assessed in the different studies can be found in Multimedia Appendix 3.

Discussion

Principal Findings

Psychoeducation is a key element for CLBP treatment. Psychoeducation provided in AR could offer more benefits than that in VR or with conventional methods by integrating physical and virtual elements. Additionally, psychoeducation in AR was assumed to be superior to VR due to reported VR-associated discomforts in CLBP treatment, such as cybersickness and visual discomfort. Therefore, we conducted a literature review in the first step to evaluate research on CLBP treatment through psychoeducation using conventional methods and immersive technologies in order to design a psychoeducational intervention

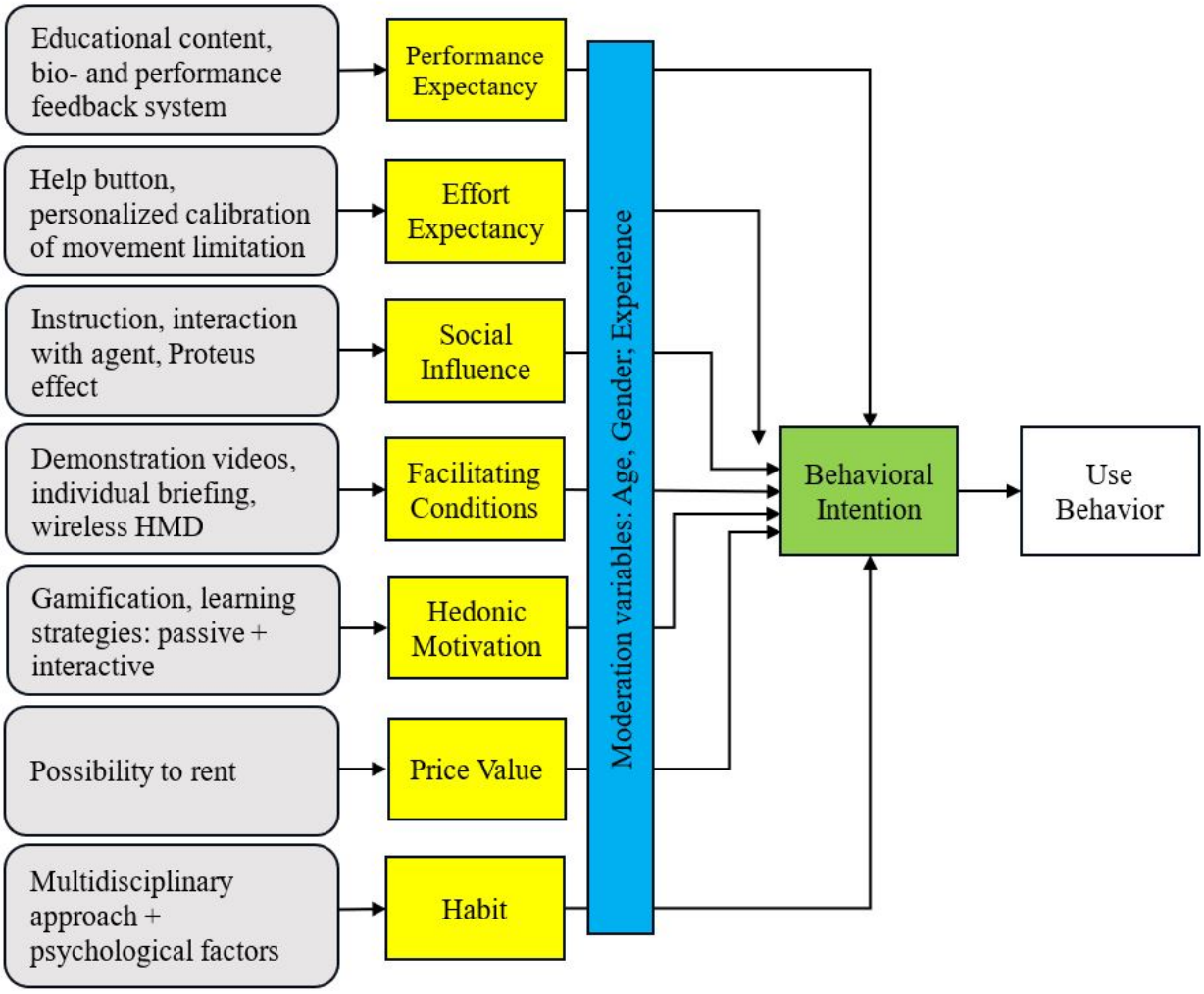
XSL•FO
RenderX

in AR for CLBP. In the second step, we applied the extracted results of the literature review to a theoretical framework, in particular, the UTAUT 2, to provide a design example for AR-based psychoeducation for CLBP.

Our findings indicate that various educational programs were combined with physiotherapy [45-48,52-54]. These studies referred to conventional methods or VR-based interventions. No relevant study with AR for CLBP treatment was found. The varying educational content encompassed pain biology [46-49], psychological education on physical activity [45,46,49,53], anxiety management [45-49,51], lifestyle modifications, daily exercises [45,49,52], coping strategies [45,52,54], pain sensitization [47,54,55], and relaxation techniques [45,49,52-55]. Passive, active [44,50,54], and gamified strategies [50,52,55] were employed alongside the promotion of intrinsic motivation [50,51]. User-friendliness [51,53], operational support [51,52,54], and application challenges [52,53] were considered important for software development. The key variables of educational CLBP interventions included physiological variables such as pain intensity [45-49,51,52,54,55] and disability level [45-47,51,52,54]; psychological variables such as kinesiophobia [46,48,51,52,54,55], pain catastrophizing [48,53-55], quality of life [51,52,55], well-being [47], self-efficacy [47,54], and depression [46,53]; and technical variables such as dropout rates [49,53] and user experience [52,53].

Our results elucidate key aspects of a useful design of a psychoeducational treatment in AR for CLBP, which does not exist to date, to the best of our knowledge. Our findings point out the relevance of the interplay of technical and psychological components, in particular, the health psychological aspects incorporating psychology of learning to foster behavior change. In the next step, the findings were applied to a theoretical framework. For this, we referred to the UTAUT 2 [8,9]. UTAUT 2 encompasses constructs such as (1) performance expectancy, (2) effort expectancy, (3) social influence, (4) facilitating conditions, (5) hedonic motivation, (6) price value, and (7) habit for intention formation as well as the moderating variables age, gender, and experience [8]. Therefore, we recommend the following design suggestions for psychoeducational interventions in AR based on UTAUT 2 for the treatment of CLBP, as exemplified in Figure 2.

Figure 2. Exemplary mapping of the extracted findings from the literature review (grey) applied to the Unified Theory of Acceptance and Use of Technology to design an artificial reality–based psychoeducation for chronic low back pain. HMD: head-mounted display.



XSL•FO
RenderX

UTAUT 2 Constructs

Performance Expectancy

It is recommended to convey psychoeducational content that demonstrates how CLBP can be positively influenced through physical activity [45,46,49,53], emotion management in kinesiophobia [45-49,51], pain-specific coping strategies [45,52,54], and stress and mindfulness techniques [45,49,52-55]. This content can be conveyed through a biofeedback system and supportive agent. As many educational measures are combined with physiotherapy [45-48,52-54], the agent can provide movement exercises, educational content, and interactive stress management techniques [54], supplemented by biofeedback. The biofeedback level and multimodal feedback of avatars promote top-down and bottom-up processes and enable associative learning [50]. To avoid discrepancies between instructions and sensory feedback, facilitate rapid corrections, and enhance user-friendliness [52], the avatar should provide immediate visual-acoustic performance feedback [50]. Biofeedback and body feedback are essential interventions for behavioral modification [17]. It is recommended to combine psychoeducation with a mindfulness-based stress reduction body scan and heart rate biofeedback [51], wherein biofeedback demonstrates progress and enables gamification elements [50,52].

Effort Expectancy

Software design should incorporate a gaming environment with in-game level settings to facilitate personalized calibration of movement limitations and body size, particularly for therapeutic activities [51,52]. The system should enable therapists to intervene through a help button or emergency assistance when patients experience pain, anxiety, or perform exercises incorrectly [51]. Frequent feedback for incorrect exercise execution should be avoided, as it may be perceived as disruptive [52].

Social Influence

With regard to social influence, an AR intervention should be accompanied by an agent that conveys pain-specific knowledge through lectures as passive knowledge transfer [44] or through support in psychological interactions, such as stress management exercises [54]. An agent can monitor a patient's body movements to integrate the phenomenon of "virtual body ownership" into the body image or utilize the analgesic effect of the Proteus effect to promote behaviors in the real world [50]. This aligns with BCT's recommendations for behavior change, wherein instruction, repetition, and demonstration of behavior have positive effects on physical activity that persist for up to 6 months [50,51]. Therefore, we propose to increase the intention to use by incorporating an agent in an AR intervention, with both passive and interactive roles.

Facilitating Conditions

A user-friendly system requires personalized instructions, demonstration videos, and individual briefings. For safety considerations, a spacious environment and wireless head-mounted displays are essential to mitigate the risk of falls. An assistant should be present to support the system. Automated breaks are crucial to prevent overexertion and exacerbation of pain, thereby automatically balancing the active therapy and relaxation periods. The VR headset should be designed for independent removal and application [51].

Hedonic Motivation

Passive learning strategies tend to yield superior outcomes in pain and functionality, whereas active methods can elicit significant improvements in knowledge [44]. Gamification demonstrates a motivation-enhancing effect through active and interactive patient engagement [50,52], for instance, through the interactive development of pain-specific knowledge in the Reducept program, where users enter their own brains and shoot with laser guns or connect points [55], or through movement exercises on a simulated farm [52]. A feedback system or biofeedback could be integrated, as outlined in the variable "performance expectancy" of UTAUT 2 and should be phrased positively as praise to increase motivation [51]. Praise as a social reward can occur through interaction with an agent, as described in the variable "social influence" [18]. In CBT, praise serves as positive reinforcement to promote adaptive behaviors and cognitions corresponding to positive CBT, which incorporates positive psychology and solution-focused brief therapy into a cognitive-behavioral context [57]. Gamification in an AR-based intervention enables the implementation of BCTs [17] by creating a material incentive such as within the framework of a game in an AR application [43].

Price Value

The headsets required for the interventions should be provided or loaned rather than purchased [50].

Habit

Educational programs for CLBP should incorporate a multidisciplinary approach that encompasses both physiological and psychological pain while promoting behavioral modifications such as regular physical activity [49]. Conventional recommendations for behavioral change emphasize repetition as crucial for habit formation [50]. For long-term interventions aimed at behavioral modification, theories addressing the intention-behavior gap and behavioral automaticity in physical activity should be considered, such as the Affective Reflective Theory of Physical Inactivity and Exercise [58] or the Physical Activity Adoption and Maintenance model [59].

Strengths and Limitations

This scoping review gives an overview of the most important educational content, elements of psychological training, interactive design forms, and relevant pain psychological variables for developing CLBP interventions in AR. It offers a substantiated basis for a theory-based development of a psychoeducational treatment in AR. Thus, this study provides a framework for the theory-driven extraction of hypotheses for future AR research in CLBP treatment. One limitation of this review encompasses the exclusion of certain sport science and physiotherapy databases (eg, SPORTDiscus) and the restriction to studies published in German and English, potentially omitting relevant publications. Further, the distinction between CLBP and chronic nonspecific low back pain in the included studies

XSL•FO
RenderX

was often imprecise. The distinction might be relevant for the intervention design, which was neglected in our analysis.

Recommendations for Research

First, the theoretically proposed design of AR-based psychoeducation for CLBP should be realized in future research. Second, an evaluation of the feasibility and user experience is needed. Third, the therapeutic efficacy of the psychoeducational content in AR must be demonstrated in a clinical evaluation study with patients with CLBP. As no studies on the psychometric properties of measurements in AR are known, psychometric assessments must be tested for measurement equivalence.

Conclusions

For the development of a framework for an AR-based psychoeducational intervention in CLBP, the combination of theories of acceptance and use of technologies with insights from health psychological behavior change theories appears to be of great importance. An example for a theory-based design of psychoeducation in AR for CLBP is proposed and discussed. Our results offer a substantiated basis for a theory-based development of psychoeducational treatment in AR.

Acknowledgments

This work was funded in part by the Ministry for Science and Health of Rhineland-Palatinate as part of the research training group (Forschungskolleg) “Immersive Extended Reality for Physical Activity and Health” (XR-PATH). The publication was funded or supported by the Open Access Fund of Universität Trier and by the German Research Foundation (DFG).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Contents of the education, duration, and physiotherapy involvement.
[DOCX File , 18 KB - ijmr_v14i1e59611_app1.docx]

Multimedia Appendix 2
Summary of the study characteristics, results, and conclusions of the 12 included studies.
[DOCX File , 19 KB - ijmr_v14i1e59611_app2.docx]

Multimedia Appendix 3
Summary of the pain-specific measuring outcomes and instruments.
[DOCX File , 17 KB - ijmr_v14i1e59611_app3.docx]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) extension for scoping reviews checklist.
[DOCX File , 108 KB - ijmr_v14i1e59611_app4.docx]

References

1. Alsufiany M, Lohman E, Daher N, Gang G, Shallan A, Jaber H. Non-specific chronic low back pain and physical activity: A comparison of postural control and hip muscle isometric strength: A cross-sectional study. Medicine (Baltimore) 2020 Jan;99(5):e18544 [FREE Full text] [doi: 10.1097/MD.0000000000018544] [Medline: 32000363]
2. Schreijenberg M, Koes BW, Lin CC. Guideline recommendations on the pharmacological management of non-specific low back pain in primary care - is there a need to change? Expert Rev Clin Pharmacol 2019 Feb;12(2):145-157. [doi: 10.1080/17512433.2019.1565992] [Medline: 30618319]
3. Faucett J. Chronic low back pain: early interventions. Annu Rev Nurs Res 1999 Jan 01;17(1):155-182. [doi: 10.1891/0739-6686.17.1.155]
4. Dobbels F, De Bleser L, Berben L, Kristanto P, Dupont L, Nevens F, et al. Efficacy of a medication adherence enhancing intervention in transplantation: The MAESTRO-Tx trial. J Heart Lung Transplant 2017 May;36(5):499-508. [doi: 10.1016/j.healun.2017.01.007] [Medline: 28162931]
5. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, COST B13 Working Group on Guidelines for Chronic Low Back Pain. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006 Mar;15 Suppl 2(Suppl 2):S192-S300 [FREE Full text] [doi: 10.1007/s00586-006-1072-1] [Medline: 16550448]
6. Schäfer R, Niederer D, Levenig C, et al. Effectiveness and mechanisms of a multimodal treatment for low back pain: A pragmatic matched cohort study. MedRxiv. Preprint posted online on February 2, 2023 2023. [doi: 10.1101/2022.09.28.22280380]

XSL•FO
RenderX

7. Alfuth M, Lages A. Chronischer Rückenschmerz. CME 2018 Jun 13;15(6):9-24. [doi: 10.1007/s11298-018-6424-8]
8. Frizziero A, Pellizzon G, Vittadini F, Bigliardi D, Costantino C. Efficacy of core stability in non-specific chronic low back pain. JFMK 2021 Apr 22;6(2):37. [doi: 10.3390/jfmk6020037]
9. Mueller J, Niederer D. Dose-response-relationship of stabilisation exercises in patients with chronic non-specific low back pain: a systematic review with meta-regression. Sci Rep 2020 Oct 09;10(1):16921 [FREE Full text] [doi: 10.1038/s41598-020-73954-9] [Medline: 33037280]
10. Minobes-Molina E, Nogués MR, Giralt M, Casajuana C, de Souza DLB, Jerez-Roig J, et al. Effectiveness of specific stabilization exercise compared with traditional trunk exercise in women with non-specific low back pain: a pilot randomized controlled trial. PeerJ 2020;8:e10304 [FREE Full text] [doi: 10.7717/peerj.10304] [Medline: 33312766]
11. Sutanto D, Ho RST, Poon ETC, Yang Y, Wong SHS. Effects of different trunk training methods for chronic low back pain: a meta-analysis. IJERPH 2022 Mar 01;19(5):2863. [doi: 10.3390/ijerph19052863]
12. Isaikin AI, Nasonova TI, Mukhametzyanova AK. Emotional disorders and their therapy in chronic low back pain. RJTAO 2022 Oct 25;14(5):90-95. [doi: 10.14412/2074-2711-2022-5-90-95]
13. Comachio J, Magalhães MO, Campos Carvalho E Silva APDM, Marques AP. A cross-sectional study of associations between kinesiophobia, pain, disability, and quality of life in patients with chronic low back pain. Adv Rheumatol 2018 Jun 22;58(1):8. [doi: 10.1186/s42358-018-0011-2] [Medline: 30657061]
14. Williams FMK, Elgaeva EE, Freidin MB, Zaytseva OO, Aulchenko YS, Tsepilov YA, et al. Causal effects of psychosocial factors on chronic back pain: a bidirectional Mendelian randomisation study. Eur Spine J 2022 Jul;31(7):1906-1915 [FREE Full text] [doi: 10.1007/s00586-022-07263-2] [Medline: 35662366]
15. Silva-Smith AL, Hanson CL, Neubeck L, Rowat A, McHale S. Physical activity interventions framed by the health action process approach for adults with long-term conditions: a scoping review. Int J Behav Med 2024 Dec;31(6):987-1017. [doi: 10.1007/s12529-024-10305-2] [Medline: 39009797]
16. Venkatesh V, Thong JYL, Xu X. Consumer acceptance and use of information technology: extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly 2012;36(1):157. [doi: 10.2307/41410412]
17. Or C. Towards an integrated model: task-technology fit in unified theory of acceptance and use of technology 2 in education contexts. JALT 2023 Jan 20;6(1):1-2. [doi: 10.37074/jalt.2023.6.1.8]
18. Barbosa RAP, Martins AVS, Da Silva IPA, Costa LA, Oliveira RA, De Souza HC. Acceptance and use of a virtual learning environment (VLE): structural equations modeling of the unified theory of acceptance and use of technology. Int J Innov Educ Res 2020 Apr 01;8(4):237-244. [doi: 10.31686/ijier.vol8.iss4.2285]
19. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
20. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. Correction: A web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2019 Feb 06;21(2):e12015 [FREE Full text] [doi: 10.2196/12015] [Medline: 30726188]
21. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, 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 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
22. Milgram P, Kishino F. A taxonomy of mixed reality visual displays. IEICE Trans Inf Syst. 1994. URL: https://cs.gmu.edu/~zduric/cs499/Readings/r76JBo-Milgram_IEICE_1994.pdf [accessed 2024-01-13]
23. Skarbez R, Smith M, Whitton MC. Revisiting Milgram and Kishino's reality-virtuality continuum. Front Virtual Real 2021 Mar 24;2:1. [doi: 10.3389/frvir.2021.647997]
24. Mallari B, Spaeth EK, Goh H, Boyd BS. Virtual reality as an analgesic for acute and chronic pain in adults: a systematic review and meta-analysis. J Pain Res 2019;12:2053-2085 [FREE Full text] [doi: 10.2147/JPR.S200498] [Medline: 31308733]
25. Awan A. Enhancing user experience: exploring mobile augmented reality experiences. VFAST Transactions on Software Engineering 2024:121-132. [doi: 10.21015/vtse.v12i1.1741]
26. Trujillo MS, Alvarez AF, Nguyen L, Petros J. Embodiment in virtual reality for the treatment of chronic low back pain: a case series. J Pain Res 2020;13:3131-3137 [FREE Full text] [doi: 10.2147/JPR.S275312] [Medline: 33269003]
27. Tack C. Virtual reality and chronic low back pain. Disabil Rehabil Assist Technol 2021 Aug;16(6):637-645. [doi: 10.1080/17483107.2019.1688399] [Medline: 31746250]
28. Hennessy RW, Rumble D, Christian M, Brown DA, Trost Z. A graded exposure, locomotion-enabled virtual reality app during walking and reaching for individuals with chronic low back pain: cohort gaming design. JMIR Serious Games 2020 Aug 10;8(3):e17799 [FREE Full text] [doi: 10.2196/17799] [Medline: 32773381]
29. Garcia LM, Darnall BD, Krishnamurthy P, Mackey IG, Sackman J, Louis RG, et al. Correction: self-administered behavioral skills-based at-home virtual reality therapy for chronic low back pain: protocol for a randomized controlled trial. JMIR Res Protoc 2021 Feb 12;10(2):e27652 [FREE Full text] [doi: 10.2196/27652] [Medline: 33577465]
30. Garcia LM, Birckhead BJ, Krishnamurthy P, Sackman J, Mackey IG, Louis RG, et al. An 8-week self-administered at-home behavioral skills-based virtual reality program for chronic low back pain: double-blind, randomized, placebo-controlled

XSL•FO
RenderX

trial conducted during COVID-19. J Med Internet Res 2021 Feb 22;23(2):e26292 [FREE Full text] [doi: 10.2196/26292] [Medline: 33484240]
31. Li R, Li Y, Kong Y, Li H, Hu D, Fu C, et al. Virtual reality-based training in chronic low back pain: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2024 Feb 26;26:e45406 [FREE Full text] [doi: 10.2196/45406] [Medline: 38407948]
32. Nagpal AS, Raghunandan A, Tata F, Kibler D, McGeary D. Virtual reality in the management of chronic low back pain: a scoping review. Front Pain Res (Lausanne) 2022;3:856935 [FREE Full text] [doi: 10.3389/fpain.2022.856935] [Medline: 35295809]
33. Wu YF, Kim EY. Users' perceptions of technological features in augmented reality (AR) and virtual reality (VR) in fashion retailing: a qualitative content analysis. Mobile Information Systems 2022 Aug 4;2022:1-13. [doi: 10.1155/2022/3080280]
34. Sala N. IGI Global. Virtual reality, augmented reality, and mixed reality in education 2020:48-73. [doi: 10.4018/978-1-7998-4960-5.ch003]
35. Tamilmani K, Rana NP, Wamba SF, Dwivedi R. The extended Unified Theory of Acceptance and Use of Technology (UTAUT2): A systematic literature review and theory evaluation. International Journal of Information Management 2021 Apr;57:102269. [doi: 10.1016/j.ijinfomgt.2020.102269]
36. 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 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
37. Vernooij RW, Sanabria AJ, Solà I, Alonso-Coello P, Martínez García L. Guidance for updating clinical practice guidelines: a systematic review of methodological handbooks. Implement Sci 2014 Jan 02;9:3 [FREE Full text] [doi: 10.1186/1748-5908-9-3] [Medline: 24383701]
38. von Elm E, Schreiber G, Haupt CC. [Not Available]. Z Evid Fortbild Qual Gesundhwes 2019 Jun;143:1-7. [doi: 10.1016/j.zefq.2019.05.004] [Medline: 31296451]
39. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616] [Medline: 21410846]
40. Khalil H, Peters MD, Tricco AC, Pollock D, Alexander L, McInerney P, et al. Conducting high quality scoping reviews-challenges and solutions. J Clin Epidemiol 2021 Feb;130:156-160. [doi: 10.1016/j.jclinepi.2020.10.009] [Medline: 33122034]
41. Mayring P, Fenzl T. Qualitative Inhaltsanalyse. In: Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer; 2019.
42. Nicol V, Verdaguer C, Daste C, Bisseriex H, Lefèvre-Colau MM, Rannou F, et al. Chronic low back pain: a narrative review of recent international guidelines for diagnosis and conservative treatment. J Clin Med 2023 Feb 20;12(4):1685 [FREE Full text] [doi: 10.3390/jcm12041685] [Medline: 36836220]
43. WHO Guideline Recommendations on Digital Interventions for Health System Strengthening. Geneva: World Health Organization; 2019.
44. Salazar-Méndez J, Cuyul-Vásquez I, Ponce-Fuentes F, Guzmán-Muñoz E, Núñez-Cortés R, Huysmans E, et al. Pain neuroscience education for patients with chronic pain: A scoping review from teaching-learning strategies, educational level, and cultural perspective. Patient Educ Couns 2024 Jun;123:108201. [doi: 10.1016/j.pec.2024.108201] [Medline: 38387389]
45. Ferlito R, Blatti C, Lucenti L, Boscarino U, Sapienza M, Pavone V, et al. Pain education in the management of patients with chronic low back pain: a systematic review. JFMK 2022 Sep 26;7(4):74. [doi: 10.3390/jfmk7040074]
46. Rim M, Leila R, Aicha BT, Olfa S, Meriem H, Ines L, et al. Efficiency of associating therapeutic patient education with rehabilitation in the management of chronic low back pain: a randomized controlled trial. Korean J Fam Med 2022 Nov;43(6):367-373 [FREE Full text] [doi: 10.4082/kjfm.21.0223] [Medline: 36444121]
47. Sidiq M, Muzaffar T, Janakiraman B, Masoodi S, Vasanthi RK, Ramachandran A, et al. Effects of pain education on disability, pain, quality of life, and self-efficacy in chronic low back pain: A randomized controlled trial. PLoS One 2024;19(5):e0294302 [FREE Full text] [doi: 10.1371/journal.pone.0294302] [Medline: 38805446]
48. Tomás-Rodríguez MI, Asensio-García MDR, García RN, Delicado-Miralles M, Sánchez SH, Segura-Heras JV. Short- and medium-term effects of a single session of pain neuroscience education on pain and psychological factors in patients with chronic low back pain. A single-blind randomized clinical trial. Eur J Pain 2024 Nov;28(10):1841-1854. [doi: 10.1002/ejp.4700] [Medline: 39017623]
49. Janik F, Fabre C, Seichepine AL, Masquelier B, Barbier F, Toulotte C. Middle-term effects of education programme in chronic low back pain patients to an adherence to physical activity: A randomized controlled trial. Patient Education and Counseling 2024 Feb;119:108081. [doi: 10.1016/j.pec.2023.108081]
50. Lindner S, Latoschik M, Rittner H. [Use of virtual reality as a component of acute and chronic pain treatment]. Anasthesiol Intensivmed Notfallmed Schmerzther 2020 Sep;55(9):549-561. [doi: 10.1055/a-1022-3038] [Medline: 32916738]
51. Stamm O, Dahms R, Müller-Werdan U. Virtual reality in pain therapy: a requirements analysis for older adults with chronic back pain. J Neuroeng Rehabil 2020 Sep 29;17(1):129 [FREE Full text] [doi: 10.1186/s12984-020-00753-8] [Medline: 32993678]

XSL•FO
RenderX

52. Stamm O, Dahms R, Reithinger N, Ruß A, Müller-Werdan U. Virtual reality exergame for supplementing multimodal pain therapy in older adults with chronic back pain: a randomized controlled pilot study. Virtual Real 2022;26(4):1291-1305 [FREE Full text] [doi: 10.1007/s10055-022-00629-3] [Medline: 35194374]
53. Brown L, DiCenso-Fleming T, Ensign T, Boyd AJ, Monaghan G, Binder DS. Chronic pain education delivered with a virtual reality headset in outpatient physical therapy clinics: a multi-site exploratory trial. Am J Transl Res 2023;15(5):3500-3510 [FREE Full text] [Medline: 37303618]
54. McConnell R, Lane E, Webb G, LaPeze D, Grillo H, Fritz J. A multicenter feasibility randomized controlled trial using a virtual reality application of pain neuroscience education for adults with chronic low back pain. Ann Med 2024 Dec;56(1):2311846 [FREE Full text] [doi: 10.1080/07853890.2024.2311846] [Medline: 38354690]
55. de Vries FS, van Dongen RTM, Bertens D. Pain education and pain management skills in virtual reality in the treatment of chronic low back pain: A multiple baseline single-case experimental design. Behav Res Ther 2023 Mar;162:104257. [doi: 10.1016/j.brat.2023.104257] [Medline: 36731183]
56. Butler D, Moseley G. Explain pain. In: 2nd ed. Australia: Noigroup Publications; 2013.
57. Bannink F, Geschwind N. Positive CBT: Individual and Group Treatment Protocols for Positive Cognitive Behavioral Therapy. Göttingen, Germany: Hogrefe; May 2021.
58. Brand R, Ekkekakis P. Affective–Reflective Theory of physical inactivity and exercise. Ger J Exerc Sport Res 2017 Nov 14;48(1):48-58. [doi: 10.1007/s12662-017-0477-9]
59. Jekauc D, Gürdere C, Englert C, Strobach T, Bottesi G, Bray S, et al. The contribution and interplay of implicit and explicit processes on physical activity behavior: empirical testing of the physical activity adoption and maintenance (PAAM) model. BMC Public Health 2024 May 06;24(1):1239 [FREE Full text] [doi: 10.1186/s12889-024-18589-5] [Medline: 38711051]

Abbreviations

AR: augmented reality
BCT: behavior change technique
CBT: cognitive behavioral therapy
CLBP: chronic low back pain
PICO: Population, Intervention, Comparison, Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
UTAUT 2: Unified Theory of Acceptance and Use of Technology 2
VR: virtual reality

Edited by M Yousufuddin; submitted 17.04.24; peer-reviewed by A Werner, Z Lu, L Pérez-Gisbert; comments to author 30.07.24; revised version received 07.11.24; accepted 10.11.24; published 20.01.25.

Please cite as:
Conen R, Mueller S, Tibubos AN
Integration of Conventional and Virtual Reality Approaches in Augmented Reality for Theory-Based Psychoeducational Intervention Design for Chronic Low Back Pain: Scoping Review
Interact J Med Res 2025;14:e59611
URL: https://www.i-jmr.org/2025/1/e59611
doi:10.2196/59611
PMID:

©Robin Conen, Steffen Mueller, Ana Nanette Tibubos. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 20.01.2025. 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

Using an Electronic Immunization Registry (Aplikasi Sehat IndonesiaKu) in Indonesia: Cross-Sectional Study

Dewi Nur Aisyah[1,2,3,4], PhD; Astri Utami[2], BSc; Fauziah Mauly Rahman[2,5], BSc; Nathasya Humaira Adriani[2], BSc; Fiqi Fitransyah[2], BSc; M Thoriqul Aziz Endryantoro[2], BSc; Prima Yosephine Hutapea[6], MPH; Gertrudis Tandy[6], MPH; Logan Manikam[1,3], PhD; Zisis Kozlakidis[7], PhD

[1]Department of Epidemiology and Public Health, Institute of Epidemiology and Health Care, University College London, London, United Kingdom

[2]Digital Transformation Office, Ministry of Health Republic of Indonesia, Jakarta, Indonesia

[3]Aceso Global Health Consultants Pte Limited, Singapore, Singapore

[4]Department of Public Health, Monash University, Tangerang, Indonesia

[5]School of Computer Science, Faculty of Engineering, University of Sydney, Sydney, Australia

[6]Directorate of Immunization Management, Ministry of Health Republic of Indonesia, Jakarta, Indonesia

[7]International Agency for Research on Cancer, World Health Organization, Lyon, France

Corresponding Author:
Logan Manikam, PhD
Department of Epidemiology and Public Health
Institute of Epidemiology and Health Care
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 2076792000
Email: logan.manikam.10@ucl.ac.uk

Abstract

Background: Electronic immunization registries (EIRs) are being increasingly used in low- and middle-income countries. In 2022, Indonesia's Ministry of Health introduced its first EIR, named Aplikasi Sehat IndonesiaKu (ASIK), as part of a comprehensive nationwide immunization program. This marked a conversion from traditional paper-based immunization reports to digital routine records encompassing a network of 10,000 primary health centers (*puskesmas*).

Objective: This paper provides an overview of the use of ASIK as the first EIR in Indonesia. It describes the coverage of the nationwide immunization program (Bulan Imunisasi Anak Nasional) using ASIK data and assesses the implementation challenges associated with the adoption of the EIR in the context of Indonesia.

Methods: Data were collected from primary care health workers' submitted reports using ASIK. The data were reported in real time, analyzed, and presented using a structured dashboard. Data on ASIK use were collected from the ASIK website. A quantitative assessment was conducted through a cross-sectional survey between September 2022 and October 2022. A set of questionnaires was used to collect feedback from ASIK users.

Results: A total of 93.5% (9708/10,382) of public health centers, 93.5% (6478/6928) of subdistricts, and 97.5% (501/514) of districts and cities in 34 provinces reported immunization data using ASIK. With >21 million data points recorded, the national coverage for immunization campaigns for measles-rubella; oral polio vaccine; inactivated polio vaccine; and diphtheria, pertussis, tetanus, hepatitis B, and *Haemophilus influenzae* type B vaccine were 50.1% (18,301,057/36,497,694), 36.2% (938,623/2,595,240), 30.7% (1,276,668/4,158,289), and 40.2% (1,371,104/3,407,900), respectively. The quantitative survey showed that, generally, users had a good understanding of ASIK as the EIR (650/809, 80.3%), 61.7% (489/793) of the users expressed that the user interface and user experience were overall good but could still be improved, 54% (422/781) of users expressed that the ASIK variable fit their needs yet could be improved further, and 59.1% (463/784) of users observed sporadic system interference. Challenges faced during the implementation of ASIK included a heavy workload burden for health workers, inadequate access to the internet at some places, system integration and readiness, and dual reporting using the paper-based format.

Conclusions: The EIR is beneficial and helpful for monitoring vaccination coverage. Implementation and adoption of ASIK as Indonesia's first EIR still faces challenges related to human resources and digital infrastructure as the country transitions from

XSL•FO
RenderX

paper-based reports to electronic or digital immunization reports. Continuous improvement, collaboration, and monitoring efforts are crucial to encourage the use of the EIR in Indonesia.

(Interact J Med Res 2025;14:e53849) doi:10.2196/53849

KEYWORDS

immunization; registry; digital; puskesmas; public health center; mobile app

Introduction

Background

Immunization is an important aspect of maintaining the health of individuals and society as a whole. It is a life-saving measure and a highly cost-effective public health intervention, an indispensable element of primary health care that shields children from avoidable mortality, illnesses, and disabilities caused by highly transmissible diseases [1]. Routine immunization plays an important role in reducing the incidence of various infectious diseases such as diphtheria, *Haemophilus influenzae* type B, measles, polio, rubella, and tetanus [2]. Located off the coast of Southeast Asia as the largest archipelagic country with 5 main islands, Indonesia has the fourth largest population in the world at >277 million [3]. This large population potentially poses the risk of rapid disease transmission if they are not protected by immunization. To ensure universal access to immunization services, vaccines must be distributed to geographically isolated regions as well as culturally or socially distinct populations, including hard-to-reach groups such as displaced individuals; migrants; and those impacted by conflict, political instability, and natural disasters [4]. Ensuring comprehensive immunization coverage for the entire population is of utmost importance yet challenging [3].

In recent years, Indonesia has faced challenges in achieving complete basic immunization, with 2020 and 2021 rates reaching 84.2%, falling short of the coverage targets of 92.9% and 93.6%, respectively [5,6]. Catch-up immunization for measles-rubella (MR) in toddlers also had declining coverage, standing at 65.3% (of a target of 76.4%) in 2020 and 58.5% (of a target of 81%) in 2021 [7]. Even with the increase in complete basic immunization coverage in 2022 (94.6%), immunization for preventable diseases remained low in low-coverage areas [5]. The COVID-19 pandemic significantly contributed to this decline as it led to disruptions in community immunization services, hindering the further attainment of herd immunity [5,6,8,9]. To address this issue, the Indonesian Ministry of Health (MoH) introduced the nationwide catch-up immunization campaign, which was called Bulan Imunisasi Anak Nasional (BIAN), in 2022 focusing on improving immunization coverage.

The MoH periodically conducts these nationwide catch-up immunization campaigns [7]. In 2022, Indonesia focused on conducting BIANs for two main activities: (1) supplementary immunization for MR and (2) catch-up immunization for children aged <5 years [7]. This focus was chosen based on several recommendations [7]. First, the National Committee for Measles-Rubella and Congenital Rubella Syndrome Elimination of Indonesia recommended to accelerate the achievement of MR and congenital rubella syndrome elimination targets in Indonesia by strengthening routine MR immunization with doses 1 and 2, aiming for a minimum coverage of 95%. Second, the National Immunization Technical Advisory Group recommended implementing inactivated polio vaccine (IPV) catch-up immunization for infants and children who had missed their scheduled IPV immunization to close immunity gaps and provide protection against the polio virus type 2. Third, the expert committee on diphtheria control recommended catch-up immunization efforts to close immunity gaps, especially among children aged <5 years (toddlers), and outbreak response immunization in areas experiencing outbreaks using appropriate vaccines for the target age group.

The BIAN was held initially for 1 month in May 2022 for the Sumatra, Kalimantan, Sulawesi, Nusa Tenggara, Maluku, and Papua islands and in August 2022 for the Java and Bali islands [7]. A total of 4 vaccines (oral polio vaccine [OPV]; IPV; MR; and diphtheria, pertussis, tetanus, hepatitis B, and *H influenzae* type B [DPT-HB-Hib]) were selected to support Indonesia's commitment to control global diseases, such as polio eradication, elimination of MR and congenital rubella syndrome and hepatitis B, control of diphtheria, reduction of the incidence of tuberculosis, and elimination of maternal and neonatal tetanus [7]. During the implementation of this BIAN, for the first time, Indonesia used an electronic immunization registry (EIR) capturing individual-level records [5] and replacing paper-based records as the latter were shown previously to include inaccuracies due to inexact, incomplete data and late reporting [10-12].

Objectives

EIRs as part of an immunization information system are a tool designed to provide information on immunization programs' target populations. By definition, EIRs include information that facilitates active search to identify individuals' vaccination data and vaccine history and provide support for determining which individuals need to be vaccinated and monitoring of immunization dropouts [13]. The 2 main databases needed for EIRs are demographic data and vaccination event data, which aim to identify the vaccine recipient and the vaccination event itself [13]. EIRs have been widely used in high-income countries as such digital reporting tools can improve reporting timeliness, precision, and overall performance [14]. Indonesia continues to improve its digital health sector and has launched the digital health transformation blueprint, aiming to create an integrated and sustainable health system [15]. As part of this effort, the government embarked on digitizing immunization reporting, shifting from a manual collection of aggregate data to individual digital records using the Aplikasi Sehat IndonesiaKu (ASIK). This paper aims to provide an overview of ASIK and its use and specifically evaluate ASIK implementation during the nationwide immunization campaign (BIAN) period in Indonesia.

XSL•FO
RenderX

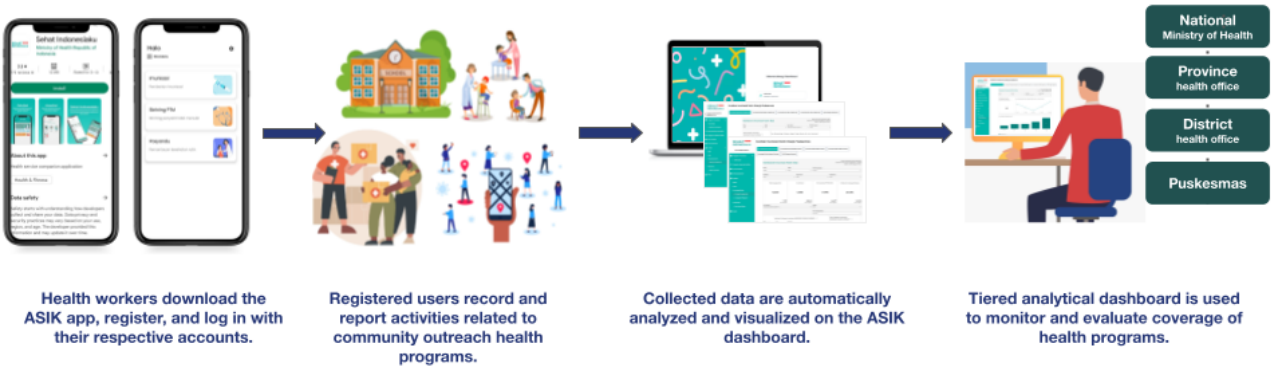
Methods

ASIK Overview

ASIK is an integrated application system developed by the MoH of Indonesia for health care workers to capture health care services provided for communities outside public health centers (*puskesmas*) [16,17]. This application is designed to help health care workers in primary care settings record patient data, report health care services provided for each individual, input routine health screening data, and monitor patient conditions [16,17].

The ASIK mobile app is available in an Android version and can be accessed through the Google Play Store. Targeted users of ASIK are health care workers in primary care settings. Health care workers need to register their account, as shown in Figure 1. Once registered and verified, health care workers can report any health care services provided outside the public health centers using ASIK to record individual data. Data reported through the ASIK mobile app are collected and automatically presented in an analytical dashboard provided in a web version of ASIK [17]. This web version of ASIK can only be accessed using a predetermined username and password for public health centers, district health offices, and provincial health offices to monitor and evaluate the coverage of health care services in their respective areas. The data presented in the dashboard are automatically synchronized from the public health centers level up to the national level. A simplified flow for the use of ASIK is shown in Figure 1.

Figure 1. Health service program reporting scheme using Aplikasi Sehat IndonesiaKu (ASIK). Puskesmas: Indonesian for "primary health center.".



ASIK was designed in line with elements of ideal EIRs according to the Pan American Health Organization's definition [18]. These elements encompass (1) inclusion of all persons at birth or as early as possible; (2) unique ID (national ID; biometrics or birth registration; or a unique combination of variables such as names, date or place of birth, and parental names or IDs); (3) information about each person, including information on geographical area of residence; (4) information about the vaccines administered, dates, and provider; (5) aggregation of data by geographical level as required; (6) timely individualized follow-up of vaccination schedule; (7) data entry as close to vaccination as possible (time and place); and (8) data security and protection of patient confidentiality.

The elements of ideal EIRs in ASIK were incorporated as follows: (1) allowing health workers to record individuals as soon as they are born with no restriction on age; (2) allowing health workers to record unique identification using the Indonesia resident registration number and a combination of name and date of birth as mandatory variables; (3) allowing health workers to record information about each individual using place of birth and current area of residence, which includes province name, city and district name, subdistrict name, village name, and detailed address; (4) allowing individuals to record information about vaccination, which includes vaccine name, vaccine batch number, date of vaccination, and location of vaccination; (5) a tiered dashboard that displays aggregated and structured data in spatial format from the national level to the village level; and (6) protection of patient confidentiality by ensuring a verification process during user registration and a limited access to a dashboard that displays detailed individual vaccination information (this access is designed specifically for the public health centers).

BIAN Immunization Reporting Using ASIK

The MoH launched and tested ASIK during the BIAN in May 2022 [19]. The BIAN 2022 immunization campaign was held in various mass public facilities such as *posyandu* (community-based health care service in Indonesia where local volunteers [cadres], midwives, and health workers provide essential maternal and child health services, including immunization); public health centers; subsidiary public health centers; hospitals; clinics; immunization service posts in schools, Islamic schools (*madrasas*), and Islamic boarding schools (*pesantrens*); and other strategic community-gathering places that could be turned into immunization posts [7]. Through ASIK, health care workers were able to record immunization data individually [16]. All immunization records from ASIK were monitored by the immunization program coordinators in public health centers in near real time.

Several databases served as the sources of ASIK data, including the public health center registry (Aplikasi Registrasi Puskesmas), the MoH Data and Information Center database, the Electronic Logistic Management and Information System (SMILE), and the Ministry of Home Affairs database. The Aplikasi Registrasi Puskesmas provides information regarding the location of immunization services, including names and locations of public health centers along with registered codes for each health facility. ASIK used data from the MoH Data and Information Center to acquire information on the number of targeted immunization participants for every district and city. Integration

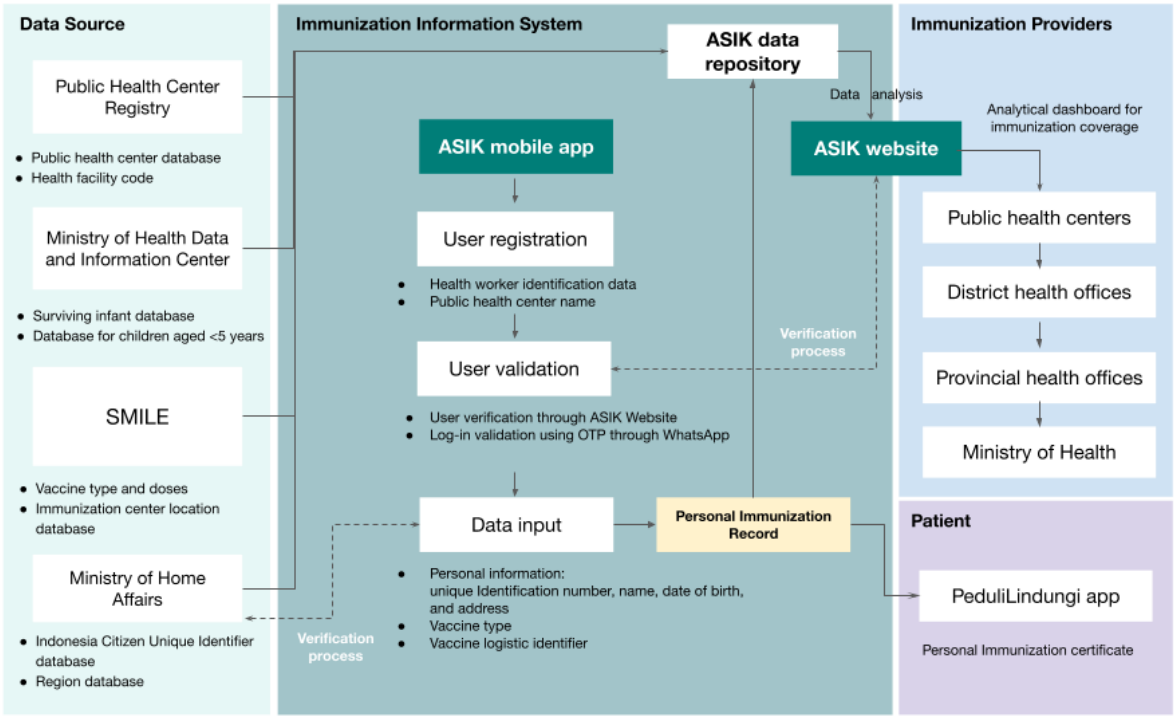
XSL•FO
RenderX

with the SMILE database allowed ASIK to access data on the types of vaccines available, the quantity of vaccine doses, and the distribution locations. During the process of recording individual data, an interconnected application programming interface with the Ministry of Home Affairs ensured data validation for citizen identification numbers and domicile information.

Each record generated through ASIK is stored in a separate database and then automatically sent to the ASIK dashboard for data analysis. Processed data are distributed to various beneficiaries, such as health facilities, district and provincial health offices, and citizens, through different platforms. In addition, individuals who have been vaccinated can receive their immunization history data through the PeduliLindungi app, which was rebranded in 2023 as SATUSEHAT Mobile [20]. Figure 2 illustrates the entire data flow within ASIK.

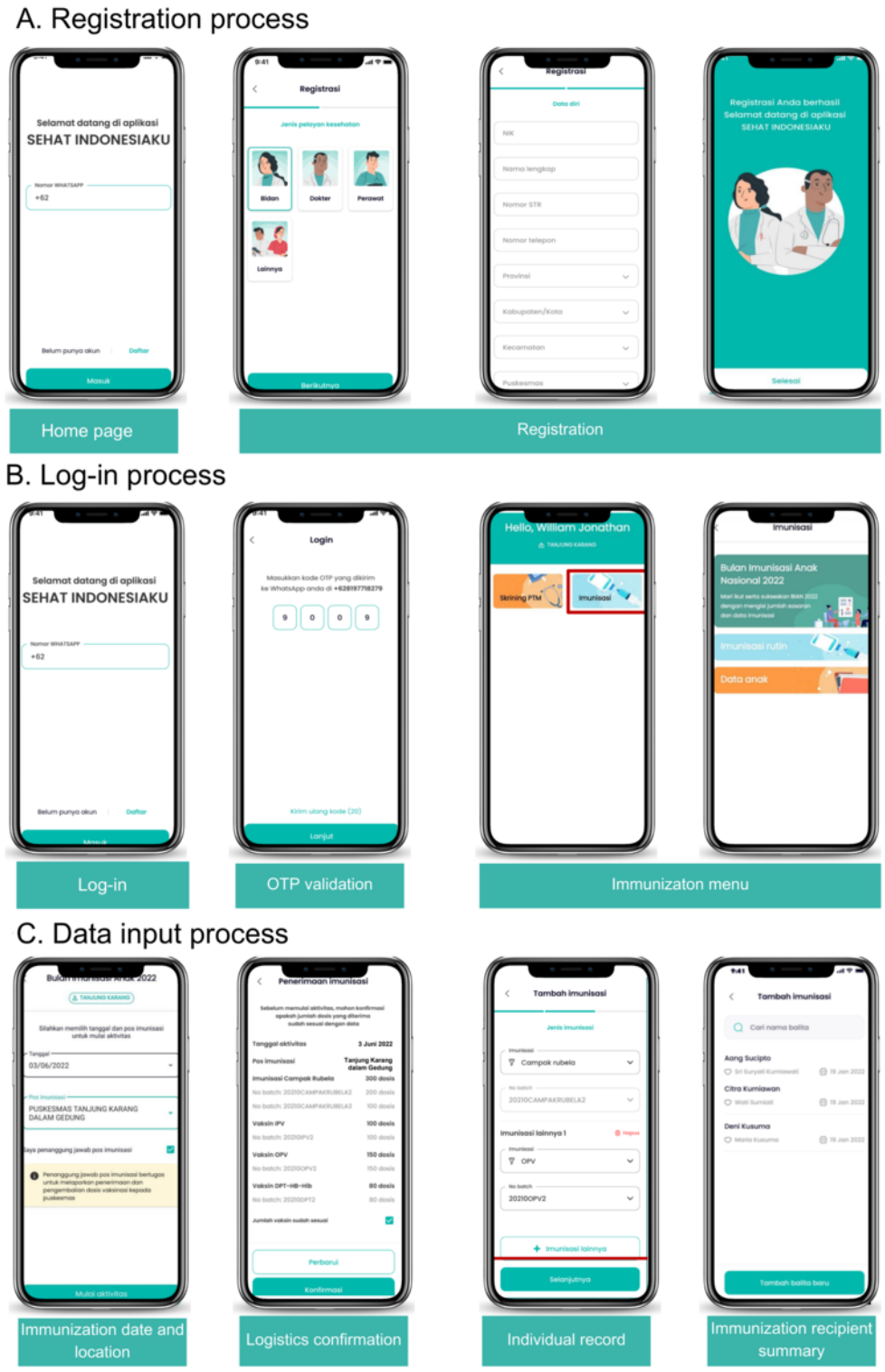
The immunization data collection process in ASIK (Figure 3) comprises the steps outlined in Textbox 1.

Figure 2. Immunization data flow in Aplikasi Sehat IndonesiaKu (ASIK). OTP: one-time password; SMILE: Electronic Logistic Management and Information System.



XSL•FO
RenderX

Figure 3. Aplikasi Sehat IndonesiaKu user interface for (A) registration process, (B) log-in process, and (C) data input process. OTP: one-time password.



XSL•FO
RenderX

Textbox 1. Immunization data collection process in Aplikasi Sehat IndonesiaKu.

User registration

Vaccinators are required to undergo registration to report immunization services using ASIK. The registration process involves providing personal information, specifying the work area, and selecting a specific public health center. Upon completing the registration form, users need to input their phone numbers to receive a one-time password for app log-in. The one-time password is sent through WhatsApp to each registered user to ensure system and user security.

Immunization data input

Vaccinators can record immunization data at designated immunization centers. Before selecting the immunization centers, the pharmaceutical officers must report the vaccine stock distribution to the specific immunization center through the SMILE. Once the logistic stock has been reported, vaccinators can choose a specific immunization center and the date of the immunization service. They can then input immunization data, including the names and addresses of recipients and the type of vaccine administered to each individual.

ASIK analytic dashboard

Immunization records captured by ASIK are integrated into an analytic dashboard. Public health center immunization coordinators can monitor the results of immunization services through this dashboard. The analytic dashboard also provides a spatial representation of the data, organized at provincial, district, and city levels. Thus, immunization data from the public health centers level to the national level can be monitored in near real time, with updates every 2 hours.

Individual immunization record

Parents can access their children's immunization records through the PeduliLindungi mobile app. Originally developed as a mass surveillance tool during the COVID-19 pandemic, PeduliLindungi initially presented COVID-19 vaccine certificates and later extended its functionality to include vaccine certificates from ASIK. Following the revocation of pandemic status, PeduliLindungi was further enhanced to encompass personal medical records. As a result, in 2023, the app was rebranded as a citizen health app named SATUSEHAT Mobile.

Data Collection

Data entry is initiated by selecting the BIAN from the immunization menu within ASIK. In total, 2 distinct roles can be chosen, namely, immunization program coordinator or immunization officer. The immunization program coordinator (ie, the individual in charge of the immunization posts) validates the available vaccine doses for each immunization post to activate the individual data report feature. This involves verifying data related to logistical distribution by reporting the number of vaccine distributions and the return of vaccination doses. On the other hand, immunization officers or vaccinators focus on reporting vaccination data at the individual level.

To record individual data, either the 16 digits of the national identification number or a combination of the full name and date of birth are entered to locate the appropriate personal health record. Consequently, a list of relevant individuals appears, allowing the users to select the relevant data or create a new identification record if the data are not available. Upon completing the identification data, vaccinators proceed to enter information about the administered vaccine type. To minimize error, the system was designed to display only the list of available vaccines suitable for the individual's age. For instance, for children aged 0 to 3 months, the system will not show the IPV in ASIK as the IPV has a minimum age requirement of 4 months. The system automatically records the results of the provided immunization services. It was recommended that vaccinators input the data immediately after services have been provided, although the option for backdating the data input is also available in ASIK, allowing users to submit delayed reports.

A quantitative survey was conducted from September 19, 2022, to October 21, 2022, using cross-sectional methods with a total of 1065 participants. All immunization program coordinators (ASIK users) from all public health centers, district health offices, and provincial health offices from 2 regions were invited to a web-based evaluation meeting. The first region represented a low data input into ASIK (the data gaps between manual and electronic reporting were approximately 30,000 to >100,000 immunization record data), and the second region represented the highest data input into ASIK (data gaps were of <10%). A total of 17.1% (182/1065) of the participants were provincial and district health officers, and 82.9% (883/1065) of the participants were health care workers from the public health centers. A questionnaire was used to collect feedback from the ASIK users (Multimedia Appendix 1). The survey used the Mentimeter platform, which enabled participants to provide direct answers and showed the cumulative, anonymized results directly. The questionnaire consisted of four sections: (1) overall user feedback on the system, (2) data reporting process, (3) supporting infrastructure, and (4) data completeness.

Data Analysis

Three outcomes were assessed in this study: (1) ASIK implementation coverage, (2) BIAN vaccination coverage (MR, OPV, IPV, and DPT-HB-Hib immunization), and (3) ASIK use feedback from users. ASIK implementation coverage and BIAN immunization coverage were calculated from the ASIK database. BIAN vaccination coverage was presented as percentage data; it was calculated by dividing the number of vaccinations (numerator) by the targeted population number for each vaccination (denominator) at the province level.

The results of BIAN immunization services were automatically presented in a structured dashboard that was accessible for public health centers, district health offices, provincial health offices, and the MoH. These data can be queried using predetermined variables. For instance, the analysis includes metrics such as the coverage of children administered the MR vaccine, OPV, IPV, and DPT-HB-Hib vaccine by comparing the service outcome rates (ie, the number of vaccinated children compared to the target numbers). The target estimates were

XSL•FO
RenderX

calculated based on the 2022 Health Development Program Target Population data [7]. Data analysis of the results of immunization services was also presented in a spatial format at the provincial, district, and city levels. Trends in service outcomes were depicted using time-series graphs. Moreover, comparative results of immunization coverage were generated (eg, by sex and age groups).

ASIK implementation coverage was measured by calculating how many public health centers used ASIK for the nationwide immunization program. The coverage was presented as percentage data, calculated by dividing the number of public health centers that reported BIAN immunization services using ASIK (numerator) by the total number of public health centers in Indonesia (denominator). The information was presented in the dashboard using the monitoring and evaluation feature. This feature enables relevant stakeholders to monitor the compliance of public health centers in reporting BIAN immunization services via ASIK. The monitoring and evaluation feature also provides information that represents the coverage of subdistricts, districts, and cities that have submitted their reports through ASIK. In addition, the dashboard generates a list of public health centers that have yet to submit their reports through ASIK. As a result, the provincial, district, and city health offices can effectively plan and implement a tiered monitoring and evaluation process.

ASIK use was assessed using a quantitative survey to obtain users' feedback. The outcome of the quantitative survey was to measure users' feedback on ASIK and how data collection using the digital tool was carried out, including (1) overall user feedback on the system (user understanding, user acceptance based on user interface [UI] and user experience [UX] and the variables presented, and user-friendliness of the system), (2) data reporting implementation (number of staff members for data input, flow of data reporting, and time to input data into ASIK), (3) infrastructure (mobile phone ownership and internet access), and (4) data completeness (availability of data on vaccinated individuals, comprising ID number, full name, date of birth, and gender). Outcomes 2, 3, and 4 were also explored to understand the real-life situations that potentially affected the use of an electronic reporting system in public health centers. The questions in the survey were developed as closed-ended questions, which allowed respondents to choose from a set of predefined answers quickly. The quantitative survey was analyzed using descriptive analysis. The data analysis was conducted using SPSS (version 20; IBM Corp) to describe users' feedback regarding the mobile app use in 2 different regions.

Ethical Considerations

The data collected for this paper do not require ethics approval in accordance with the Indonesian MoH *National Guideline and Standards for Ethical Research and Development in Health* guideline (2017, Chapter IIIB, point 2) [21] as no individual-level data are presented. The immunization analysis data are in anonymized, aggregate format and were handled following the aforementioned guidelines. For the survey, this research was conducted under the ethics approval of the International Agency for Research on Cancer Ethics Committee (22–37). Anonymous participation followed verbal informed consent, which was considered an indication of consent for obtaining and analyzing the data. There was no compensation provided for participants as it is part of the MoH immunization campaign program. No individual data were presented; only aggregated anonymized data were used as per the aforementioned guidelines.

Results

BIAN Immunization Reports by Province and Vaccine Type

The BIAN was held in 2 phases. The first phase began in May 2022 and ran until the end of July 2022, targeting 27 million children across 26 provinces in the Sumatra, Kalimantan, Sulawesi, Nusa Tenggara, Maluku, and Papua islands. The second phase took place in August 2022 in the Java and Bali islands (7 provinces), aiming for roughly 10 million children. The data analysis presented in this paper depicts the BIAN reports between May 2022 and December 2022. Different areas had distinct vaccination targets for each vaccine (Multimedia Appendix 2).

According to the data records in ASIK, there were 21.89 million BIAN immunization services recorded, with details on 18.3 million for the MR vaccine (18,301,057/36,497,694, 50.1%), 938,000 for the OPV (938,623/2,595,240, 36.2%), 1.2 million for the IPV (1,276,668/4,158,289, 30.7%), and 1.37 million for the DPT-HB-Hib vaccine (1,371,104/3,407,900, 40.2%).

The highest coverage of MR immunization was attained by Banten, reaching 96.8% (901,465/931,740), followed by Central Java (1,996,112/2,069,562, 96.5%) and East Java (1,797,337/2,352,409, 76.4%). In contrast, Aceh reported the lowest coverage (93,127/1,444,337, 6.4%), followed by Papua (72,368/792,523, 9.1%) and Central Sulawesi (104,316/708,642, 14.7%).

For OPV immunization, 2 provinces surpassed 100% coverage, namely, Lampung (49,833/37,827, 131.7%) and Banten (111,255/105,771, 105.2%), followed by Central Java (115,771/131,724, 87.9%). A significant contrast emerged in OPV coverage, with Papua (951/118,227, 0.8%), East Kalimantan (658/26,294, 2.5%), and Central Kalimantan (2857/58,389, 4.9%) showing the lowest coverage.

For IPV immunization, Banten reached the highest coverage (163,826/173,701, 94.3%), followed by Bali (1254/1426, 87.9%). In contrast, Papua (1814/167,678, 1.1%), East Kalimantan (747/67,544, 1.1%), and Central Kalimantan (1892/95,243, 2%) recorded the lowest IPV coverage.

For DPT-HB-Hib immunization, 2 provinces surpassed 100% coverage, namely, Banten (188,804/172,619, 109.4%) and Bali (3733/3620, 103.1%), followed by Lampung (64,043/71,504, 89.6%). Conversely, Papua (1294/194,510, 0.7%), Aceh (7282/184,475, 3.9%), and Central Sulawesi (2041/43,691, 4.7%) recorded the lowest coverage (Multimedia Appendix 3).

In summary, significant regional disparities in immunization coverage are apparent based on ASIK data. The Java and Bali islands exhibited the highest coverage across all vaccine types (MR: 7,544,342/9,435,097, 80%; OPV: 637,763/1,102,238,

XSL•FO
RenderX

57.9%; IPV: 927,925/1,747,414, 53.1%; DPT-HB-Hib: 1,087,268/1,759,356, 61.8%), whereas the Sulawesi island had the lowest coverage for 3 vaccine types (MR: 1,480,388/4,270,493, 34.7%; IPV: 25,270/406,208, 6.2%; DPT-HB-Hib: 28,745/315,809, 9.1%). Low coverage was also observed in the Papua, Maluku, and Nusa Tenggara islands for all vaccine types.

Using the specified BIAN 2022 campaign targets as benchmarks [7], we compared the immunization coverage achieved in each province with their predetermined targets. The immunization coverage report (Multimedia Appendix 4) illustrates that only 2 provinces, Banten and Lampung, successfully attained the coverage goal for MR immunization. In parallel, 4 provinces (Bali, Banten, Central Java, and Lampung) exceeded the target for OPV immunization. Similarly, the immunization target for IPV was met only by Bali, Banten, and Central Java. The DPT-HB-Hib immunization target was accomplished only by Bali, Banten, Central Java, and Lampung. These data are in line with past reports in which immunization targets were often unmet in the context of Indonesia, although the current technological implementation allowed for more granular data and for a quicker comparison at a national scale.

Digital Versus Paper-Based Recording of BIAN Implementation

Although the implementation of the BIAN and the release of ASIK happened simultaneously, the technical implementation guideline document for BIAN stated that manual or paper-based reports are still possible using a specified format [7]. Hence, all provinces involved also recorded data using a paper-based format. This manual data collection focuses on reporting aggregate numbers of immunization services, such as the total number of children aged 12 to 59 months receiving DPT-HB-Hib immunization. The data presented in the following paragraphs reflect the paper-based recording of the same campaign used to compare the digital versus paper-based recording of BIAN implementation.

Thus, according to the official letter from the MoH regarding the achievements of the BIAN phase I and phase II in 2022 and using the data from the paper-based records alone, the immunization coverage in Indonesia for MR, OPV, IPV, and DPT-HB-Hib was 72.7% (26,529,596/36,497,639), 54.2% (1,330,928/2,454,340), 45.8% (1,842,869/4,024,564), and 61% (2,011,057/3,294,942), respectively [22] (Multimedia Appendix 3).

The highest coverage of MR immunization was attained by East Java, reaching 100.6% (2,365,820/2,352,401), followed by Jakarta (710,757/715,786, 99.3%) and Banten (918,323/931,739, 98.6%). Conversely, Aceh reported the lowest coverage for MR (280,017/1,444,335, 19.4%), followed by Papua (307,466/792,523, 38.8%) and Riau (880,867/1,913,263, 46%).

In terms of OPV immunization, 3 provinces surpassed 100% coverage, namely, Bali (786/732, 107.4%), Central Java (113,325/106,194, 106.7%), and Banten (107,570/105,085, 102.4%). A significant contrast emerged in OPV coverage, with West Nusa Tenggara (10,224/129,413, 7.9%), West Kalimantan (6556/78,005, 8.4%), and Central Kalimantan (5472/58,389, 9.4%) showing the lowest coverage.

Regarding IPV immunization, Yogyakarta attained the highest coverage at 92.2% (400/434), followed by West Java (611,346/664,213, 92%) and East Java (195,047/212,128, 91.9%). Conversely, Papua (5521/167,678, 3.3%), East Kalimantan (2744/67,544, 4.1%), and Central Kalimantan (4378/95,243, 4.6%) recorded the lowest IPV coverage.

For DPT-HB-Hib immunization, South Sumatra led with the highest coverage (36,117/36,397, 99.2%), followed by Bali (3606/3639, 99.1%) and Central Java (293,073/297,231, 98.6%). In contrast, Papua (9676/194,510, 5%), Aceh (25,323/195,289, 13%), and West Kalimantan (8510/79,635, 10.7%) reported the lowest DPT-HB-Hib coverage (Multimedia Appendix 3).

On the basis of the table in the Multimedia Appendix 3, there were gaps between manual reporting data as compared to the digital recording via ASIK.

For MR immunization, the average reporting gap was 24.4% ±3.72% (95% CI: 20.68%-28.12%); the highest gap was observed in Maluku (47.4%), followed by South Sulawesi (47.3%) and Central Sulawesi (47.3%). The greatest reporting alignment was observed in Central Java (1.5%), followed by Banten (1.9%) and South Kalimantan (11.6%).

For OPV immunization, the average gap was 18.3% ±7.26% (95% CI: 11.04%-25.56%); the highest discrepancy was observed in East Java (70.4%), followed by Gorontalo (60.6%) and the Riau islands (53.6%). The greatest alignment was observed in Lampung (−39.7%), followed by Banten (−2.8%) and Jakarta (−1.9%).

For IPV immunization, the average gap was 16.1% ±6.42% (95% CI: 9.68%-22.52%); the highest data discrepancy was observed in the Special Region of Yogyakarta (77.4%), followed by East Java (70.8%) and the Riau islands (56.5%). The closest reporting alignment was observed in Jakarta (−1.8%), followed by West Nusa Tenggara (1.6%) and Papua (2.2%).

For DPT-HB-Hib immunization, the average gap was 20.7% ±7.23% (95% CI: 13.47%-27.93%); the highest data reporting discrepancy was observed in the Riau islands (82.4%), followed by East Java (71.6%) and Maluku (58.2%). The greatest reporting alignment was observed in Banten (−12.2%), followed by Bali (−4%) and Lampung (−0.4%).

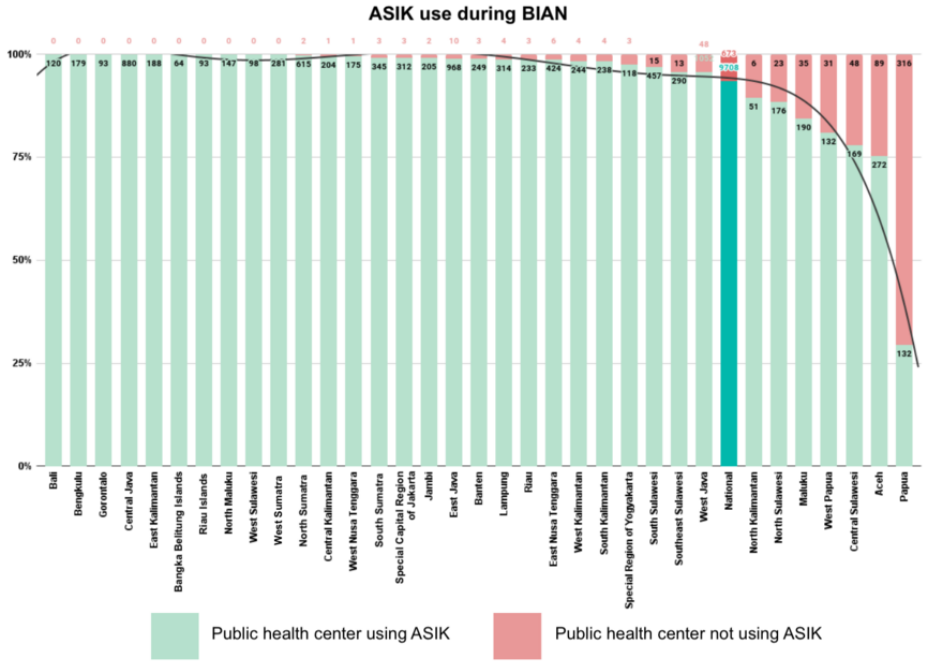
ASIK Use and Coverage

To understand the aforementioned discrepancies, 2 further aspects were considered: ASIK use and coverage and user feedback on ASIK use. The results will be presented in the following paragraphs. Regarding ASIK coverage, a total of 93.5% (9708/10,382) of public health centers, 93.5% (6478/6928) of subdistricts, and 97.5% (501/514) of districts and cities in 34 provinces reported BIAN activities using ASIK. On the basis of the coverage as measured at the province level, there were 10 provinces where 100% of their public health centers reported BIAN activities using ASIK, namely, Bali, Bengkulu, Gorontalo, Central Java, East Kalimantan, Bangka Belitung Islands, Riau Islands, North Maluku, West Sulawesi, and West Sumatra, whereas the provinces with the lowest

XSL•FO
RenderX

number of public health centers reporting using ASIK were Papua (132/448, 29.5%), Aceh (272/361, 75.3%), and Central Sulawesi (169/217, 77.9%). Detailed information on other provinces can be found in Figure 4.

Figure 4. Aplikasi Sehat IndonesiaKu (ASIK) use across 34 provinces during the Bulan Imunisasi Anak Nasional (BIAN) implementation.



User Feedback on ASIK Use

In parallel to understanding the coverage and use of ASIK, a quantitative survey was conducted to obtain users' feedback on the system, including explorations of UI and UX, reporting variables, and system interference. A total of 1065 questionnaires were received, although not all respondents answered all questions. The results of the quantitative survey are presented in this section and summarized in Table 1. In general, users exhibited a good understanding of ASIK as the EIR (650/809, 80.3%). For UI and UX, the highest feedback was that it was overall good but could still be improved (489/793, 61.7%), with 35.3% (280/793) of respondents expressing that the UI and UX were very good and 3% (24/793) saying that they were hard to understand. Regarding the required variables for immunization data input in ASIK, 54% (422/781) of the respondents expressed that the variables fit their needs but could still be improved, 39.6% (309/781) expressed that the variables fit their needs, and 6.4% (50/781) of users expressed that there were too many variables. In terms of system interferences, 59.1% (463/784) of users expressed that they sometimes happened, and 24.5% (192/784) expressed that they often happened. The most prevalent system interferences were (1) inability to find individual data in the system (281/779, 36.1%), (2) server errors (217/779, 27.9%), and (3) problems with logistic stock confirmation (135/779, 17.3%).

On the basis of the users' feedback, 44% (382/869) stated that there were 2 to 3 staff members that helped with the data reporting process. However, the number of staff members who helped with data input varied as users in region 2 (high data collection) had >5 staff members for data reporting and users in region 1 (low data collection) only had 2 to 3 staff members to do so. Regarding the flow of data reporting, 33.4% (291/872) of users exhibited a combination of recording the immunization data on paper or on Microsoft Excel first and directly inputting them into ASIK. Furthermore, more users preferred the input of immunization data at public health centers right after the immunization activity was completed (364/1089, 33.4%), followed by inputting the data within 24 hours after the immunization services (203/1089, 18.6%). However, 22.3% (181/813) of users in region 1 input the immunization data 2 weeks after the immunization activity, whereas in region 2, a total of 45.3% (125/276) of their users preferred to input the data right after the activity was completed. Regarding the data reporting process, each region had different characteristics for each variable: number of staff members, data reporting flow, and time to input the data. Region 2 showed a relatively higher number of staff members, more direct flow of data reporting, and timelier input of data into the system. As previously mentioned, region 2 had the highest data collection, comprising 28 cities that did not have any gaps in data collection between ASIK and manual reporting.

XSL•FO
RenderX

Table 1. Quantitative exploration of the use of Aplikasi Sehat IndonesiaKu (ASIK; N=1065).

Characteristic	Region 1/ (n=800), n/ (%)	Region 2/ (n=265), n/ (%)	Total, n/ (%)
Overall user feedback on the system			
System understanding			
Limited understanding	135/598 (22.6)	24/211 (11.4)	159/809 (19.7)
Good understanding	463/598 (77.4)	187/211 (88.6)	650/809 (80.3)
User interface and user experience			
Very good	200/587 (34.1)	80/206 (38.8)	280/793 (35.3)
Overall good but can still be improved	366/587 (62.4)	123/206 (59.7)	489/793 (61.7)
Hard to understand	21/587 (3.6)	3/206 (1.5)	24/793 (3)
Variables			
Fit with needs	213/576 (37)	96/205 (46.8)	309/781 (39.6)
Fit with needs but can still be improved	321/576 (55.7)	101/205 (49.3)	422/781 (54)
Too many variables	42/576 (7.3)	8/205 (3.9)	50/781 (6.4)
System interference			
Seldom	18/579 (3.1)	3/205 (1.5)	21/784 (2.7)
Rarely happened	58/579 (10)	36/205 (17.6)	94/784 (12)
Sometimes happened	326/579 (56.3)	137/205 (66.8)	463/784 (59.1)
Often happened	163/579 (28.2)	29/205 (14.1)	192/784 (24.5)
Always happened	14/579 (2.4)	0/205 (0)	14/784 (1.8)
Data reporting process			
Number of staff members for data input			
1 person	95/645 (14.7)	8/224 (3.6)	103/869 (11.9)
2-3 people	311/645 (48.2)	71/224 (31.7)	382/869 (44)
4-5 people	98/645 (15.2)	43/224 (19.2)	141/869 (16.2)
>5 people	141/645 (21.9)	102/224 (45.5)	243/869 (28)
Flow of data reporting			
(1) Recording the data on paper or manual book report before inputting them into ASIK	153/647 (23.6)	35/225 (15.6)	188/872 (21.6)
(2) Recording the data in Microsoft Excel sheet before inputting them into ASIK	119/647 (18.4)	20/225 (8.9)	139/872 (15.9)
(3) Directly inputting the data into ASIK	51/647 (7.9)	46/225 (20.4)	97/872 (11.1)
Combination of 1 and 2	131/647 (20.2)	26/225 (11.6)	157/872 (18)
Combination of 1, 2, and 3	193/647 (29.8)	98/225 (43.6)	291/872 (33.4)
Time to input data into ASIK			
Directly at the vaccination center	78/813 (9.6)	60/276 (21.7)	138/1089 (12.7)
Directly after the activity at the public health centers	239/813 (29.4)	125/276 (45.3)	364/1089 (33.4)
Input the data within a day at home	161/813 (19.8)	42/276 (15.2)	203/1089 (18.6)
Input the data within 3 days of manual data collection	86/813 (10.6)	14/276 (5.1)	100/1089 (9.2)
Input the data within 7 days of manual data collection	68/813 (8.4)	14/276 (5.1)	82/1089 (7.5)
Input the data ≥2 weeks after manual data collection	181/813 (22.3)	21/276 (7.6)	202/1089 (18.5)
Existing infrastructure			
Android mobile phone ownership			
Yes	694/711 (97.6)	229/241 (95)	923/952 (97)[y]

XSL•FO
RenderX

Characteristic	Region 1/ (n=800), n/ (%)	Region 2/ (n=265), n/ (%)	Total, n/ (%)
No	17/711 (2.4)	12/241 (5)	29/952 (3)[y]
Internet access			
Available at all places	390/710 (54.9)	163/240 (67.9)	553/950 (58.2)
Only available at the public health centers	78/710 (11)	14/240 (5.8)	92/950 (9.7)
Limited availability	242/710 (34.1)	63/240 (26.3)	305/950 (32.1)
Data completeness			
Data availability			
Individual data were complete	157/637 (24.6)	125/219 (57.1)	282/856 (32.9)
Individual data were incomplete	480/637 (75.4)	94/219 (42.9)	574/856 (67.1)

Regarding the existing infrastructure, 97% (923/952) of users had Android mobile phones. A total of 58.2% (553/950) of respondents expressed that internet access was available at all places, 32.1% (305/950) expressed that internet availability was limited in several places, and 9.7% (92/950) expressed that the internet was only available at public health centers (and not available in immunization posts). The data suggested that the internet was more accessible in region 2 compared to region 1.

Another constraint found during data reporting was data completeness of the targeted participants. A total of 67.1% (574/856) of users expressed that the individual data were incomplete, especially on unique identifier (ID number), which was only available for up to 79% of the collected data points; full name, which was available for up to 95% of individual data points; birth date, which was available for up to 95% of the collected data points; and gender, which was available for up to 95% of the collected data points.

Most respondents (618/1065, 58%) indicated several constraints that hindered the use of ASIK and were experienced intermittently or too frequently, including (1) challenges in retrieving individual data, (2) server errors, and (3) logistic stock validation. The first challenge that arose in the early development of the system is the lack of interconnectedness with the civil registration database. During the preliminary development phase of ASIK, the absence of established data sharing agreements between the Indonesian MoH and the Ministry of Home Affairs prevented users from efficiently engaging in data reporting using ASIK.

Interconnection between ASIK and the civil registration database will help health care workers automatically retrieve individual demographic information such as gender, date of birth, and address by inputting the 16 digits of the national ID. During BIAN implementation, ASIK was still not connected to the civil registration database owned by the Ministry of Home Affairs; thus, health care workers needed to input the individual information manually.

The second challenge also stemmed from the early development phase of the system, when the initial server capacity was constrained, leading to suboptimal application performance. System errors that were mentioned by the respondents happened occasionally throughout the recording process. They sometimes happened due to personal data input, vaccine type data input, or the process of saving data. They potentially happened due to limited server performance caused by the high number of users inputting data at the same time.

The third challenge pertains to logistic stock management. In contrast to the first challenge, ASIK was strategically integrated with the logistic system (SMILE) during its development phase, effectively establishing a linkage between immunization and logistic data. However, the compliance and adherence to using the SMILE were variable. Consequently, the comprehensive implementation of the SMILE for monitoring logistic stocks in the context of the BIAN 2022 campaign was not universal across all public health centers. This posed an obstacle in the use of ASIK given that logistic confirmation is a prerequisite for inputting immunization data into ASIK. Respondents were hindered from inputting the immunization data into ASIK if the pharmacy staff had not input the logistic information into the SMILE. This created some delays in the data recording process during BIAN 2022.

These challenges have been addressed, and the system has been enhanced over time, particularly during the latter stage of the BIAN 2022 campaign. The system was integrated with the civil registration database, the server capacity was upgraded, and refinements were implemented in the integration between ASIK and the SMILE. As a result, the efficacy of ASIK had a notable improvement during the latter phase of the BIAN 2022 campaign. Nevertheless, many respondents gave positive responses for the use of ASIK as the Indonesia EIR. The positive response correlated with the convenience of ASIK as an EIR.

Discussion

Principal Findings

The launch of ASIK is a part of the MoH's mission to transform health care services by assisting health care workers in efficiently and comprehensively recording patient data in a single, integrated database [5,15,19]. The use of ASIK as Indonesia's first EIR was broad, covering 34 out of 38 provinces. The rate of adoption reached 93.5% (9708/10,382) of public health centers (ie, the main health care provider for the nationwide immunization campaign). Through 2 phased releases, ASIK managed to record >21 million immunizations. Although most users expressed a good understanding of the

XSL•FO
RenderX

system, several challenges persisted during the implementation phase.

EIRs are instruments proven to be cost-effective in enhancing coverage and timeliness of vaccination [23]. The use of EIRs was considered beneficial in preventing delays in vaccinating children as they help parents keep track of their children's vaccination records and stay informed of their children's vaccination schedule [23,24]. Furthermore, EIRs are helpful to identify high-risk populations; facilitate resource and activity planning, such as monitoring vaccination coverage and vaccine stock availability; and determine the overall performance of the immunization program [23,24]. During ASIK's first launch, Indonesia's MoH emphasized that, in the future, children who have received immunization will have their records digitally stored in ASIK, in which the integrated data aim to facilitate parents' access to their children's immunization data, especially to be used for school-related purposes [19].

In the technical implementation guideline document for the BIAN, recording and reporting of immunization activities are mandated to be conducted electronically through ASIK [7]. Despite electronic data input being mandatory, the data recorded in ASIK did not represent the complete immunization service during the BIAN 2022 campaign. According to the ASIK analytical dashboard, the national coverage recorded in ASIK for MR, OPV, IPV, and DPT-HB-Hib immunization was 50.1% (18,301,057/36,497,694), 36.2% (938,623/2,595,240), 30.7% (1,276,668/4,158,289), and 40.2% (1,371,104/3,407,900), respectively. In addition, there was an average discrepancy of approximately 19% between digital and manual data reporting. When compared to similar supplementary immunization efforts in previous years, the Polio National Immunization Week held in 2016 achieved an impressive coverage rate of 96.5%. MR supplementary immunization activities in 2017 exceeded expectations, reaching >100.9%, whereas in 2018, the coverage stood at approximately 73.4% [7]. Around the same time, between 2017 and 2019, several countries, including Afghanistan, Benin, Ghana, Timor-Leste, Togo, Sierra Leone, Senegal, Rwanda, Pakistan, Niger, Malawi, Lesotho, and Ethiopia, also attained MR supplementary activity coverage rates of >90% [25]. Nonetheless, it is important to consider that these coverage data were obtained from manual or paper-based immunization records and postcampaign surveys [25].

Considering that the immunization data are stored in different files segregated by health center and area, it is nearly impossible to ensure that there is no duplication of data in each record. Studies show that paper-based data reports actually create challenges such as affecting health care workers' ability to make data-driven decisions by analyzing the manual immunization reports and increasing the burden of health workers to input data both in the paper-based and digital reporting systems [26]. Furthermore, the use of paper-based aggregate data for analysis presents gaps, such as limited individual longitudinal data, and omits crucial sociodemographic information [27]. Data stored electronically enable prompt data retrieval, hence facilitating immediate reports related to disease and surveillance, which is in contrast to paper-based reports that rely on manual retrieval [13]. Dual use of EIR and paper-based immunization reports further exacerbates the challenges of data input using ASIK for health care workers. A transition from paper-based reports to paperless reports and dual entry processes are also perceived barriers and challenges to EIR implementation in other countries such as Kenya and Vietnam [26,28]. The immunization program has grown in complexity and leads to an increased need for data, and the immunization information system must be flexible to cater to those needs [13]. Paper-based reports are limited in meeting such needs as they are heavily influenced by population mobility and errors stemming from estimation of population number [13]. Hence, moving forward, it is important for relevant stakeholders to decide whether to solely use paper-based reports or onboard using ASIK for Indonesia immunization programs.

On the other hand, data captured in ASIK represent individual data, which can be linked to personal immunization history records. Individual immunization data recorded in an EIR such as ASIK offer the advantage of providing precise and timely information. This, in turn, facilitates critical activities such as monitoring vaccination compliance, mitigating dosing errors, and identifying unvaccinated individuals [29]. The significance of individual immunization records in clinical decision-making is noteworthy as they provide a comprehensive historical vaccination record and diminish the likelihood of misclassifying vaccination status [30]. Consequently, this aids in reducing instances of both under- and overimmunization in children [31]. Furthermore, integrating individual immunization data from an immunization information system into an electronic medical record or electronic health record can enhance the accuracy of vaccine safety assessments [30,32].

ASIK has achieved significant engagement, involving 93.5% (9708/10,382) of public health centers across Indonesia within a span of 7 months since its launch. This indicates that public health centers were informed about the data input process and user guide updates that come with it, although long-term adherence to data input remains to be seen. The high level of ASIK use might be supported by several aspects, such as the compatibility of the Android app in Indonesia, the user-friendliness of the system interface, and the ease of use. Considering the high level of mobile phone and internet use in Indonesia, the digitization of health service reports is happening apace [33]. In 2021, approximately 90.5% of households in Indonesia owned at least one mobile phone, covering 65.8% of the population, and approximately 62% of the population had access to the internet [34]. Indonesia is estimated to have >200 million mobile phone users and is dominated by Android operating systems (95%) [34,35]. This is in line with our findings, in which 97% (923/952) of respondents stated that they owned Android-based mobile devices. ASIK is built on an Android operating system, which is compatible with various mobile devices, allowing health care workers in Indonesia to benefit from it [33]. Existing EIRs such as ImmunizaCA in Canada, Zindagi Mehfooz in Pakistan, and DHIS2 Android Capture App in Zambia are also built on Android-based platforms [36-40].

A visually appealing or user-friendly UI and UX can prompt users to explore an app, facilitating a comprehensive understanding of its available functions, consequently influencing user acceptability [41,42]. It is essential to acknowledge that ASIK signifies a transition for health care

XSL•FO
RenderX

workers from paper-based to electronic reporting; hence, data and variables should be both useful and recognizable to the health workers as it is important to note that a familiar interface correlates with use frequency [43]. Other than that, a familiarity with workflow and a design that carefully considers health care workers' needs are crucial for effective EIR implementation [26].

Despite the users' acceptability of ASIK as an EIR, the low number of reporting in certain provinces may be due to several factors. First, primary health care workers, often responsible for numerous data collection tasks, are frequently burdened with heavy workloads, lack motivation, and view data recording as undesirable and time-consuming [44,45]. Such excessive workloads can compromise the quality of health care data reported at health facilities, leading to inaccuracies and unreliability when other tasks interfere with data entry [46,47]. Indeed, the lowest compliance was often correlated with the provinces that have the lowest ratios of primary health care staff, and thus, it is presumed that staff time availability is the strongest root cause for the observed variation in ASIK use. The workload also affects workers' opportunities to adequately familiarize themselves (digital familiarization) with any health information system they are about to use, often coupled with inadequate handover and shortage of available staff for this knowledge transfer [48,49]. In the next phase, to ensure that there is an appropriate and sustainable knowledge transfer to all end users, conducting a cascading training-of-trainers approach and establishing a multilevel technical support network is recommended [28].

The sudden explosion in the number of health-related apps in Indonesia, with >400 health-related apps developed and introduced for health care facilities, further exacerbates the burden on health workers regarding digital familiarization and data input [15]. In addition, lack of adequate monitoring by supervisors or managers and provision of technical support affects the effective implementation of data reports through an information system [48,49]. Furthermore, the BIAN implementation coincided with a massive COVID-19 preventive immunization effort, adding more tasks to primary care workers' daily workload [50,51]. The implementation of COVID-19 immunization entailed a collaborative approach involving various sectors, with health personnel sourced from governmental, military, and private practitioner sectors [52]. This stands in stark contrast to the BIAN 2022 campaign, which relied heavily on health workers stationed at public health center facilities [52]. Therefore, there was a double burden for primary care workers to reach both the target for COVID-19 immunization and the target for the BIAN 2022 campaign and report both results using different digital pathways. This further justified the need to simplify, integrate, and ensure the interoperability of the health information system in Indonesia, as mentioned in the digital transformation blueprint, and ensure that the EIR is developed to match users' needs and able to simplify the reporting workflow accordingly [15,26].

In the process of implementing digital technologies in the health care sector, certain challenges are inevitable [53]. These challenges encompass a range of factors, including but not limited to client identification and verification as well as lack of database integration [53,54]. In ASIK, the collection of the national identification number of the service recipient is a prerequisite for generating an immunization service record. Indonesia uses a 16-digit identification number, known as the resident registration number, which serves as a cornerstone of national demographic identification [55]. Current data indicate that 97.2% of Indonesian citizens already possess a unique identification number, but there are also 5.38 million individuals without unique identification numbers [55,56]. During BIAN implementation, there were many instances in which patients did not bring any personal identification tools when immunization services were provided. Consequently, users were burdened to create individual data without unique identifiers while also needing to remember that exact record for identification and retrieval in the future when the same patient visited. Lack of linkage and integration with other databases were also a common challenge, which happened in the development of EIRs in Africa [54]. It is crucial to foster awareness and establish a mutual understanding between the realms of public health and IT to facilitate EIR development [54]. Ensuring that EIRs comprehensively covered all regions across Indonesia, encompassing an estimated target population of 37 million individuals for BIAN immunization, presented a challenge in conveying this requirement effectively, especially in the beginning of the system's development. To build an interoperable EIR system, integration with the national ID database is crucial [28]. Moving forward, the MoH could establish a continuous collaboration with the Ministry of Home Affairs as the primary stakeholders and foster understanding of and commitment to the need for database integration and providing immediate legal identity to the population since birth.

To report immunization data in ASIK, having access to the internet is a prerequisite. Despite statistics indicating that approximately 62% of Indonesia's population use the internet and the country's consistent progress in improving internet connectivity, disparities persist in terms of access across different social classes and regions [34,57]. Notably, significant portions of Indonesia, particularly in the eastern regions, still lack access to the internet [58]. The distribution of infrastructure for signal reception in Indonesia reveals that >3023 villages lack signal reception infrastructure, with >934 of these villages located in Papua [34]. Given the limited internet access, it is unsurprising that the coverage of MR, OPV, IPV, and DPT-HB-Hib immunization reports using ASIK remained low in Papua. Mobile apps that need to constantly rely on internet connectivity reduce effectiveness for users such as health care workers that operate in remote regions [59]. To reduce the digital divide among regions, the government could explore the potential use of satellite-based cellular and internet connectivity for rural areas [60]. This could be done through collaboration with the Ministry of Communication and Information Technology or alongside existing internet provider entities. Other than ensuring internet connectivity, support and assistance to infrastructure and human resources is paramount, particularly in areas with low information and communications technology maturity scores [61].

This is the first study that describes the EIR in Indonesia, which replaces the manual reporting system from the MoH, including

XSL•FO
RenderX

describing its use during the BIAN. Furthermore, the quantitative survey provided an overview of the obstacles in the implementation of ASIK for the BIAN program and provided input on the features that need to be improved for future immunization data collection processes. However, a relative weighing of the factors that require future improvement was not sought and, thus, can become the subject of future research.

This study has several limitations that need to be considered. First, ASIK is only available on Android-based platforms and has not been released for iOS; thus, data reporting is limited to officers who own devices compatible with Android. Second, the quantitative survey that was administered covered a limited number of the actual ASIK users in the field, but it is expected that this number can represent feedback from the field. In addition, using Mentimeter potentially limited the feedback from some of the participants when they did not fill the questionnaire in a timely manner, and considering that the evaluation meeting was conducted on the web, there was a risk of limited control over the participants filling out the questionnaire in a timely manner. Finally, the manual data that were compared to ASIK are the result of a report from the provincial health office, which might not fully capture the recording process that took place. Nevertheless, this study provides a deeper understanding of the use of ASIK in the BIAN program as the first EIR implemented nationally by the MoH.

Conclusions

EIRs have demonstrated their cost-effectiveness in improving vaccination coverage and timeliness, offering numerous advantages in disease prevention and health care management. Adoption and continuous use of EIRs entail numerous challenges, especially in a country such as Indonesia with an enormous number of targeted immunization populations and limited infrastructure settings. Challenges observed from this study included the gap between manually reported data and electronic records due to dual data entry processes; heavy workloads for primary care health workers to carry out immunization campaigns as well as the reporting process; lack of data completeness in ID number, full name, date of birth, or gender; system interferences due to the early development phase of the system; and limited infrastructure settings to support digital connection in remote areas. Addressing these challenges and ensuring adequate support for health care workers are essential steps in enhancing the effectiveness of EIRs in improving public health outcomes.

Although faced with the aforementioned challenges, ASIK as Indonesia's first national EIR managed to reach substantial engagement. More than 93% of its targeted users adopted the system, and this was potentially made possible by ease of system use, continuous system improvement according to users' feedback, and immediate regulation support. Ease of use and user-friendly apps that simplify data input processes are essential to reinforce user acceptability. In the future, ASIK can be developed for a wider immunization program provided that improvement of features is carried out, including the preparation of relevant human resources and supporting digital infrastructure and stronger regulatory support.

Acknowledgments

The authors would like to thank the Digital Transformation Office, Ministry of Health, Republic of Indonesia, for supporting the study implementation and developing the Aplikasi Sehat IndonesiaKu integrated system; the public health centers and district and provincial health offices across Indonesia who participated in the study; and the Directorate of Immunization Management, Ministry of Health of Indonesia, who supported the study implementation. The authors did not receive funding for the creation of this manuscript. This work was supported in part by the International Agency for Research on Cancer (IARC) through training by the Biobanking and Cohort Network (Civilian Research and Development Foundation Global, award 66415). While 2 authors are affiliated with Aceso Global Health Consultants, which is a private company, they declare that this research project did not receive funding from Aceso Global Health Consultants. The company did not have a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Where authors are identified as personnel of the IARC, the authors alone are responsible for the views expressed in this paper, which do not necessarily represent the decisions, policy, or views of the IARC.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

DNA led the study design and developed the paper concept, main structure, and methodology. AU, FMR, and NHA conducted the quantitative survey and accessed and analyzed the data. DNA, AU, FMR, and NHA wrote the first draft of the paper. ZK and LM validated the study and critically revised the manuscript content. DNA and AU had primary responsibility for the final content. All authors contributed to reviewing the study data and the manuscript design and to the approval of the final manuscript.

Conflicts of Interest

LM is the director of Aceso Global Health Consultants, and DNA is a consultant of the company. However, both of them contributed to this paper on a pro bono basis. All other authors declare no conflicts of interest.

XSL•FO
RenderX

Multimedia Appendix 1
Questionnaire.
[DOCX File , 203 KB - ijmr_v14i1e53849_app1.docx]

Multimedia Appendix 2
Bulan Imunisasi Anak Nasional 2022 campaign target.
[DOCX File , 9 KB - ijmr_v14i1e53849_app2.docx]

Multimedia Appendix 3
Comparison of measles-rubella vaccine; oral polio vaccine; inactivated polio vaccine; and diphtheria, pertussis, tetanus, hepatitis B, and Haemophilus influenzae type B immunization coverage using Aplikasi Sehat IndonesiaKu data and manual report data.
[DOCX File , 14 KB - ijmr_v14i1e53849_app3.docx]

Multimedia Appendix 4
Detail of measles-rubella vaccine; oral polio vaccine; inactivated polio vaccine; and diphtheria, pertussis, tetanus, hepatitis B, and Haemophilus influenzae type immunization coverage using Aplikasi Sehat IndonesiaKuASIK data and manual report data.
[DOCX File , 23 KB - ijmr_v14i1e53849_app4.docx]

References

1. Vaccine-preventable diseases and immunization programme. World Health Organization Regional Office for Europe. 2005. URL: https://iris.who.int/bitstream/handle/10665/107665/WHO-EURO-2005-8588-48360-71799-eng.pdf?sequence=3&isAllowed=y [accessed 2023-08-05]
2. Talbird SE, Carrico J, La EM, Carias C, Marshall GS, Roberts CS, et al. Impact of routine childhood immunization in reducing vaccine-preventable diseases in the United States. Pediatrics 2022 Sep 01;150(3):e2021056013. [doi: 10.1542/peds.2021-056013] [Medline: 35821599]
3. Indonesia population. Worldmeter. URL: https://www.worldometers.info/world-population/indonesia-population [accessed 2023-08-05]
4. Immunization Agenda 2030 Partners. Immunization agenda 2030: a global strategy to leave no one behind. Vaccine 2024 Apr 08;42 Suppl 1:S5-14 [FREE Full text] [doi: 10.1016/j.vaccine.2022.11.042] [Medline: 39004466]
5. Indonesia targets low vaccination areas to tackle decline in childhood immunization: during world immunization week, Indonesia recommits to increasing immunization. United Nations International Children's Emergency Fund. 2023 May 04. URL: https://www.unicef.org/indonesia/press-releases/indonesia-targets-low-vaccination-areas-tackle-decline-childhood-immunization [accessed 2023-08-05]
6. Impact of COVID-19 on routine immunization in Indonesia: quarterly assessment using Interactive Voice Recording (IVR) technology to assess the continuity of immunization services across all 34 provinces during the COVID-19 pandemic. United Nations International Children's Emergency Fund. URL: https://www.unicef.org/indonesia/reports/impact-covid-19-routine-immunization-indonesia [accessed 2023-09-10]
7. Petunjuk Teknis Bulan Imunisasi Anak Nasional (BIAN). Kemenkes. 2022 Jun 30. URL: https://ayosehat.kemkes.go.id/petunjuk-teknis-bulan-imunisasi-anak-nasional-bian [accessed 2023-08-05]
8. Routine immunization services during the COVID-19 pandemic. World Health Organization Regional Office for the Western Pacific. 2020 Apr 13. URL: https://apps.who.int/iris/bitstream/handle/10665/331925/Routine-immunization-services-COVID-19-eng.pdf [accessed 2023-09-10]
9. Yuliyanti S, Rifqi FM, Prasetyo RL, Nurhasanah E, Priambada AB, Abbas P. Analysis of basic immunization services during the COVID-19 pandemic at public health centers in Semarang. Open Access Maced J Med Sci 2021 Nov 26;9(E):1338-1344. [doi: 10.3889/oamjms.2021.7336]
10. Lim SS, Stein DB, Charrow A, Murray CJ. Tracking progress towards universal childhood immunisation and the impact of global initiatives: a systematic analysis of three-dose diphtheria, tetanus, and pertussis immunisation coverage. The Lancet 2008 Dec 13;372(9655):2031-2046. [doi: 10.1016/s0140-6736(08)61869-3]
11. Bosch-Capblanch X, Ronveaux O, Doyle V, Remedios V, Bchir A. Accuracy and quality of immunization information systems in forty-one low income countries. Trop Med Int Health 2009 Jan;14(1):2-10 [FREE Full text] [doi: 10.1111/j.1365-3156.2008.02181.x] [Medline: 19152556]
12. Ronveaux O, Rickert D, Hadler S, Groom H, Lloyd J, Bchir A, et al. The immunization data quality audit: verifying the quality and consistency of immunization monitoring systems. Bull World Health Organ 2005 Jul;83(7):503-510 [FREE Full text] [Medline: 16175824]
13. Electronic immunization registry: practical considerations for planning, development, implementation and evaluation. Pan American Health Organization. 2017. URL: https://iris.paho.org/handle/10665.2/34865 [accessed 2023-08-11]

XSL•FO
RenderX

14. Dolan SB, Carnahan E, Shearer JC, Beylerian EN, Thompson J, Gilbert SS, et al. Redefining vaccination coverage and timeliness measures using electronic immunization registry data in low- and middle-income countries. Vaccine 2019 Mar 22;37(13):1859-1867 [FREE Full text] [doi: 10.1016/j.vaccine.2019.02.017] [Medline: 30808566]
15. Cetak biru strategi transformasi digital kesehatan 2024. Kementerian Kesehatan Republik Indonesia. 2021. URL: https://online.fliphtml5.com/egzdb/hsaa/#p=1 [accessed 2023-08-05]
16. Sehat Indonesiaku. Google Play. URL: https://play.google.com/store/apps/details?id=com.dto.sehatindonesiaku&hl=id [accessed 2023-08-04]
17. Kemenkes RI. Dashboard ASIK - Imunisasi Rutin - Saya sudah melakukan penginputan pada ASIK Mobile tapi capaian belum bertambah pada. Kemenkes. URL: https://faq.kemkes.go.id/faq/dashboard-asik-imunisasi-rutin-saya-sudah-melakukan-penginputan-pada-asik-mobile-tapi-capaian-belum-bertambah-pada-website [accessed 2025-02-01]
18. Danovaro-Holliday MC, Contreras MP, Pinto D, Molina-Aguilera IB, Miranda D, García O, et al. Assessing electronic immunization registries: the Pan American Health Organization experience. Rev Panam Salud Publica 2019;43:e28 [FREE Full text] [doi: 10.26633/RPSP.2019.28] [Medline: 31093252]
19. Rokom. Tak lagi dicatat manual, imunisasi anak akan terdata digital di Aplikasi Sehat IndonesiaKu (ASIK). Sehat Negeriku. 2022 May 18. URL: https://sehatnegeriku.kemkes.go.id/baca/umum/20220518/4039909/tak-lagi-dicatat-manual-imunisasi-anak-akan-terdata-digital-di-aplikasi-sehat-indonesiaku-asik/ [accessed 2023-08-06]
20. Rokom. Besok PeduliLindungi resmi bertransformasi menjadi SATUSEHAT mobile. Sehat Negeriku. 2023 Mar 10. URL: https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20230228/2042474/besok-pedulilindungi-resmi-bertransformasi-menjadi-satusehat-mobile/ [accessed 2025-02-01]
21. Pedoman dan standar etik penelitian dan pengembangan kesehatan nasional. Komisi Etik Penelitian Dan Pengembangan Kesehatan Nasional Kementerian Kesehatan Republik Indonesia Tahun 2017. 2017 Nov 13. URL: https://kepk.poltekkestasikmalaya.ac.id/wp-content/uploads/2018/05/2017-KEPPKN-Standar-dan-Pedoman-.pdf [accessed 2023-09-10]
22. Capaian bulan imunisasi anak nasional tahap I dan tahap I tahun 2022. Kementerian Kesehatan Republik Indonesia. 2024. URL: https://drive.google.com/file/d/1RFE2cnJUbOBDsaiw0SSJAMEgG5YCVfhJ/view?usp=sharing [accessed 2025-02-23]
23. Bloland P, MacNeil A. Defining & assessing the quality, usability, and utilization of immunization data. BMC Public Health 2019 Apr 04;19(1):380 [FREE Full text] [doi: 10.1186/s12889-019-6709-1] [Medline: 30947703]
24. El-Khatib Z, El-Halabi S, Khdeir MA, Khader YS. Children immunization app (CImA), low-cost digital solution for supporting Syrian refugees in Zaatari camp in Jordan — general description. In: Proceedings of the IEEE/ACS 16th International Conference on Computer Systems and Applications. 2019 Presented at: AICCSA 2019; November 3-7, 2019; Abu Dhabi, United Arab Emirates. [doi: 10.1109/AICCSA47632.2019.9035211]
25. Cutts FT, Danovaro-Holliday MC, Rhoda DA. Challenges in measuring supplemental immunization activity coverage among measles zero-dose children. Vaccine 2021 Mar 01;39(9):1359-1363 [FREE Full text] [doi: 10.1016/j.vaccine.2020.11.050] [Medline: 33551302]
26. Wittenauer R, Dolan SB, Njoroge A, Onyango P, Owiso G, Rabinowitz P, et al. Usability and acceptability of electronic immunization registry data entry workflows from the health care worker perspective in Siaya, Kenya (part 3): pre-post study. JMIR Form Res 2023 Mar 30;7:e39383 [FREE Full text] [doi: 10.2196/39383] [Medline: 36995755]
27. Rolnick J. Aggregate health data in the United States: steps toward a public good. Health Informatics J 2013 Jun;19(2):137-151 [FREE Full text] [doi: 10.1177/1460458212462077] [Medline: 23715213]
28. Dang H, Dao S, Carnahan E, Kawakyu N, Duong H, Nguyen T, et al. Determinants of scale-up from a small pilot to a national electronic immunization registry in Vietnam: qualitative evaluation. J Med Internet Res 2020 Sep 22;22(9):e19923 [FREE Full text] [doi: 10.2196/19923] [Medline: 32960184]
29. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. J Big Data 2019 Jun 19;6:54. [doi: 10.1186/s40537-019-0217-0]
30. Stockwell MS, Natarajan K, Ramakrishnan R, Holleran S, Forney K, Aponte A, et al. Immunization data exchange with electronic health records. Pediatrics 2016 Jun;137(6):e20154335. [doi: 10.1542/peds.2015-4335] [Medline: 27244856]
31. McGreevy S, Murray M, Montero L, Gibson C, Comfort B, Barry M, et al. Assessing the immunization information system and electronic health record interface accuracy for COVID-19 vaccinations. JAMIA Open 2023 Apr 13;6(2):ooad026 [FREE Full text] [doi: 10.1093/jamiaopen/ooad026] [Medline: 37063406]
32. Groom HC, Crane B, Naleway AL, Weintraub E, Daley MF, Wain K, et al. Monitoring vaccine safety using the vaccine safety datalink: assessing capacity to integrate data from immunization information systems. Vaccine 2022 Jan 31;40(5):752-756 [FREE Full text] [doi: 10.1016/j.vaccine.2021.12.048] [Medline: 34980508]
33. Accelerating the app economy in Indonesia: Android and Google Play's impact in Indonesia. Access Partnership. 2023 Aug. URL: https://accesspartnership.com/wp-content/uploads/2023/08/Accelerating-the-app-economy-in-indonesia.pdf [accessed 2023-09-13]
34. Statistik telekomunikasi Indonesia 2021. Badan Pusat Statistik Indonesia. 2022 Sep 7. URL: https://www.bps.go.id/publication/2022/09/07/bcc820e694c537ed3ec131b9/statistik-telekomunikasi-indonesia-2021.html [accessed 2023-08-13]
35. Smartphone users in Indonesia. Start.io. URL: https://www.start.io/audience/smartphone-users-in-indonesia [accessed 2023-08-13]

XSL•FO
RenderX

36. Hasan S, Yousuf MM, Farooq M, Marwah N, Andrabi SA, Kumar H. e-vaccine: an immunization app. In: Proceedings of the 2nd International Conference on Intelligent Engineering and Management. 2021 Presented at: ICIEM 2021; April 28-30, 2021; London, UK. [doi: 10.1109/iciem51511.2021.9445386]
37. Sullivan E, Masood T, Javed W, Bagshaw K, Ollis S, Regmi P, et al. Electronic immunization information systems: a case report of lessons learned from implementation in Pakistan. Mhealth 2020 Jul 05;6:31 [FREE Full text] [doi: 10.21037/mhealth.2020.01.07] [Medline: 32632369]
38. Wilson K, Atkinson KM, Deeks SL, Crowcroft NS. Improving vaccine registries through mobile technologies: a vision for mobile enhanced Immunization information systems. J Am Med Inform Assoc 2016 Jan;23(1):207-211 [FREE Full text] [doi: 10.1093/jamia/ocv055] [Medline: 26078414]
39. Siddiqi DA, Abdullah S, Dharma VK, Shah MT, Akhter MA, Habib A, et al. Using a low-cost, real-time electronic immunization registry in Pakistan to demonstrate utility of data for immunization programs and evidence-based decision making to achieve SDG-3: insights from analysis of Big Data on vaccines. Int J Med Inform 2021 May;149:104413. [doi: 10.1016/j.ijmedinf.2021.104413] [Medline: 33652259]
40. Seymour D, Werner L, Mwansa FD, Bulula N, Mwanyika H, Dube M, et al. Electronic immunization registries in Tanzania and Zambia: shaping a minimum viable product for scaled solutions. Front Public Health 2019 Aug 7;7:218 [FREE Full text] [doi: 10.3389/fpubh.2019.00218] [Medline: 31440494]
41. Pratama MA, Cahyadi AT. Effect of user interface and user experience on application sales. IOP Conf Ser Mater Sci Eng 2020;879:012133. [doi: 10.1088/1757-899X/879/1/012133]
42. Carnahan E, Ferriss E, Beylerian E, Mwansa FD, Bulula N, Lyimo D, et al. Determinants of facility-level use of electronic immunization registries in Tanzania and Zambia: an observational analysis. Glob Health Sci Pract 2020 Sep;8(3):488-504. [doi: 10.9745/ghsp-d-20-00134]
43. Shen Z, Xue C, Wang H. Effects of users' familiarity with the objects depicted in icons on the cognitive performance of icon identification. Iperception 2018 Jun 11;9(3):2041669518780807 [FREE Full text] [doi: 10.1177/2041669518780807] [Medline: 29977490]
44. Nicol E, Bradshaw D, Phillips T, Dudley L. Human factors affecting the quality of routinely collected data in South Africa. Stud Health Technol Inform 2013;192:788-792. [Medline: 23920665]
45. Dharmayuda AA, Wulandari LP. Analisis beban kerja dokter umum di Puskesmas kota Denpasar dengan menggunakan metode workload indicators of staffing need. Public Health Prevent Med Arch 2015 Jul;3(1):16 [FREE Full text] [doi: 10.15562/phpma.v3i1.95]
46. Hlaing T, Myint ZM. Factors affecting data quality of health management information system at township level, Bago region, Myanmar. Int J Community Med Public Health 2022 Feb 28;9(3):1298-1307. [doi: 10.18203/2394-6040.ijcmph20220686]
47. Purwanti E, Ediyono S. The relationship between heavy workload and the nurse's performance. Int J Med Sci Clin Res Stud 2022 Jun 01;2(6):445-450. [doi: 10.47191/ijmscrs/v2-i6-01]
48. Glèlè Ahanhanzo Y, Ouedraogo LT, Kpozèhouen A, Coppieters Y, Makoutodé M, Wilmet-Dramaix M. Factors associated with data quality in the routine health information system of Benin. Arch Public Health 2014 Jul 28;72(1):25 [FREE Full text] [doi: 10.1186/2049-3258-72-25] [Medline: 25114792]
49. Khobi JA, Mtebe JS, Mbelwa JT. Factors influencing district health information system usage in Sierra Leone: a study using the technology-organization-environment framework. Electron J Inf Syst Dev Ctries 2020 Nov;86:e12140. [doi: 10.1002/isd2.12140]
50. Terbaru! Ketentuan pelaksanaan vaksin booster (Per 27 Januari 2022). Satuan Tugas Penanganan COVID-19. 2022 Jan 29. URL: https://covid19.go.id/tenaga-kesehatan/terbaru-ketentuan-pelaksanaan-vaksin-booster-27-januari-2022 [accessed 2023-08-14]
51. Nugraha RR, Miranda AV, Ahmadi A, Lucero-Prisno DE. Accelerating Indonesian COVID-19 vaccination rollout: a critical task amid the second wave. Trop Med Health 2021 Sep 22;49(1):76. [doi: 10.1186/s41182-021-00367-3] [Medline: 34548099]
52. COVID-19 vaccine post introduction evaluation (cPIE): Indonesia country report. World Health Organization. 2022. URL: https://cdn.who.int/media/docs/default-source/searo/indonesia/indonesia-cpie-report.pdf [accessed 2023-08-14]
53. Classification of digital health interventions v 1.0: a shared language to describe the uses of digital technology for health. World Health Organization. 2018. URL: https://apps.who.int/iris/bitstream/handle/10665/260480/WHO-RHR-18.06-eng.pdf [accessed 2023-08-08]
54. Namageyo-Funa A, Samuel A, Bloland P, Macneil A. Considerations for the development and implementation of electronic immunization registries in Africa. Pan Afr Med J 2018;30:81. [doi: 10.11604/pamj.2018.30.81.11951]
55. Soemartono T. The dynamic of e-KTP evaluation program in DKI Jakarta. Int J Adm Sci Organ 2014;20(2):90-97. [doi: 10.20476/jbb.v20i2.3070]
56. Tangkilisan W, Sjam JM, Supriadi T, Marota R. Legal problems of the misuse of electronic identification card (E-Ktp) data in the Indonesian general election. J Posit Sch Psychol 2022;6(6):5851-5866.

XSL•FO
RenderX

57. Beyond unicorns: harnessing digital technologies for inclusion in Indonesia. World Bank Group. 2021 Jul 28. URL: https://www.worldbank.org/en/country/indonesia/publication/beyond-unicorns-harnessing-digital-technologies-for-inclusion-in-indonesia [accessed 2023-08-14]
58. Telehealth and digital inclusion in Indonesia. High-Quality Technical Assistance for Results. 2022. URL: https://archive.ids.ac.uk/heart/wp-content/uploads/2022/01/Telehealth-and-Digital-Inclusion-in-Indonesia-Executive-Summary.pdf [accessed 2023-08-14]
59. Mitgang EA, Blaya JA, Chopra M. Digital health in response to COVID-19 in low- and middle-income countries: opportunities and challenges. Glob Policy 2021 Jul;12(Suppl 6):107-109 [FREE Full text] [doi: 10.1111/1758-5899.12880] [Medline: 34230840]
60. Akhtar MD, Mathur G, Kravchenko O, Rakhra M. Connecting the unconnected: bridging the digital divide with affordable satellite through enabled smartphones. E3S Web Conf 2023 Nov 30;453:01054. [doi: 10.1051/e3sconf/202345301054]
61. Aisyah DN, Setiawan AH, Lokopessy AF, Faradiba N, Setiaji S, Manikam L, et al. The information and communication technology maturity assessment at primary health care services across 9 provinces in Indonesia: evaluation study. JMIR Med Inform 2024 Jul 18;12:e55959 [FREE Full text] [doi: 10.2196/55959] [Medline: 39037345]

Abbreviations

ASIK: Aplikasi Sehat IndonesiaKu
BIAN: Bulan Imunisasi Anak Nasional
DPT-HB-Hib: diphtheria, pertussis, tetanus, hepatitis B, and Haemophilus influenzae type B
EIR: electronic immunization registry
IPV: inactivated polio vaccine
MoH: Ministry of Health
MR: measles-rubella
OPV: oral polio vaccine
SMILE: Electronic Logistic Management and Information System
UI: user interface
UX: user experience

Edited by T de Azevedo Cardoso; submitted 21.10.23; peer-reviewed by P Mechael, A Rusibamayila, C Weerth; comments to author 08.04.24; revised version received 01.07.24; accepted 19.12.24; published 27.03.25.

Please cite as:
Aisyah DN, Utami A, Rahman FM, Adriani NH, Fitransyah F, Endryantoro MTA, Hutapea PY, Tandy G, Manikam L, Kozlakidis Z
Using an Electronic Immunization Registry (Aplikasi Sehat IndonesiaKu) in Indonesia: Cross-Sectional Study
Interact J Med Res 2025;14:e53849
URL: https://www.i-jmr.org/2025/1/e53849
doi:10.2196/53849
PMID:

©Dewi Nur Aisyah, Astri Utami, Fauziah Mauly Rahman, Nathasya Humaira Adriani, Fiqi Fitransyah, M Thoriqul Aziz Endryantoro, Prima Yosephine Hutapea, Gertrudis Tandy, Logan Manikam, Zisis Kozlakidis. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 27.03.2025. 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

Size and Composition of Caregiver Networks Who Manage Medications for Persons Living With Dementia: Cross-Sectional Analysis of the 2011-2022 National Health and Aging Trends Study

Reed WR Bratches[1], PhD; Frank Puga[1], PhD; Paul J Barr[2], PhD; Amanda N Leggett[3], PhD; Meredith Masel[4], PhD; James Nicholas Odom[1], PhD, RN; Rita Jablonski[1], PhD

[1]School of Nursing, University of Alabama at Birmingham, 1701 University Blvd, Birmingham, AL, United States

[2]Center for Technology and Behavioral Health, Dartmouth College, Hanover, NH, United States

[3]Institute of Gerontology, Wayne State University, Detroit, MI, United States

[4]School of Public & Population Health, University of Texas Medical Branch, Galveston, TX, United States

Corresponding Author:
Reed WR Bratches, PhD
School of Nursing, University of Alabama at Birmingham, 1701 University Blvd, Birmingham, AL, United States

Abstract

Background: Family caregivers commonly help manage medications taken by persons living with dementia. Recent work has highlighted the importance of caregiver networks, which are multiple caregivers managing care for a single person, on managing care for persons living with dementia, especially medication management. However, less is known about the composition of caregiver networks.

Objective: The objective of this analysis was to describe the composition of caregiver networks that manage medications, the factors associated with helping with medications within caregiver networks, and whether racial or ethnic differences exist in caregiver network composition.

Methods: This cross-sectional secondary analysis used data from the National Health and Aging Trends Study (NHATS) "other person" files from 2011 to 2022. Descriptive statistics were calculated for caregivers who were identified as helping manage medications for a person with dementia. Mixed-effect logistic regression was used to determine factors associated with helping with medications among caregiver networks, with odds ratios converted to predicted probabilities using marginal standardization. A *P* value of .05 or less was considered statistically significant. Secondary analysis was stratified by race and ethnicity due to identified cultural differences in living situation and overall caregiver network composition.

Results: A total of 15,809 caregivers were analyzed. Of those, 3048 (19.2%) managed medications for persons living with dementia. Caregiver networks that manage medications tend to include a spouse or partner and child, at least one of whom has a college degree. Every person with dementia reported at least 1 person who managed their medications. White persons with dementia had an average of 2.4 (range 1 - 9) people who managed medications, while Black or African American persons with dementia had an average of 2.8 (range 1 - 9) and Hispanic or Latino persons with dementia had an average of 2.9 (range 1 - 8) people who managed medications. Spouses were most likely to manage medications across all racial and ethnic groups. In regression modeling, female gender (predicted probability [PP] 15%, 95% CI 13%-17%; *P*<.001), Black or African American race (PP 7%, 95% CI 4%-10%; *P*<.001), and Hispanic ethnicity (PP 4%, 95% CI 1%-9%; *P*=.04) were associated with an increased probability of helping with medications.

Conclusions: The size and composition of caregiver networks that manage medications for persons living with dementia differ by race and ethnicity but typically includes at least 2 people, one of whom has a college degree. Helping with medications was more likely among non-White family caregivers, while White patients with dementia were more likely to use paid help to manage medications.

(Interact J Med Res 2025;14:e64499) doi:10.2196/64499

KEYWORDS

caregiver; dementia; disparities; cross-sectional analysis; racial; ethnic; family caregivers; logistic regression; secondary analysis; manage; management; medication; network

XSL•FO
RenderX

Introduction

Persons living with dementia frequently take multiple medications due to comorbid conditions and pharmacologic attempts to manage the behavioral symptoms of dementia [1]. Family caregivers, or spouses, children, and friends who provide care for persons living with dementia, are often responsible for medication management [2]. Approximately 54% of family caregivers manage medications in early-stage dementia, rising to 90% in later dementia stages, as executive functioning declines and persons living with dementia become more reliant on family caregivers [3,4]. There are identified disparities in dementia diagnoses and medication management, dementia being more likely among Black and Hispanic persons [5], with family caregivers from historically minoritized racial and ethnic groups helping with medications at increased rates relative to non-Hispanic White caregivers [6,7].

Medication management of person living with dementia can be challenging and stressful for family caregivers [2,3,8]. Medication management includes components such as filling prescriptions and understanding medication purposes, organizing and administering medications, monitoring for side effects, and sustaining medication regimens, which can be very time-consuming for family caregivers [9].

Most prior studies have focused on medication management by family caregivers in the dementia dyad (eg, one person living with dementia and one family caregiver), but recent literature has highlighted the importance of caregiver networks in managing the care of persons living with dementia [10]. Caregiver networks can be defined as a cohesive network of family and friends who provide support to the persons living with dementia [11]. Caregiver networks can help reduce caregiver burden by sharing stressful tasks over multiple caregivers [12], and caregiver networks can facilitate better patient care through improved caregiver functioning [13]. Caregiver networks can be influenced by cultural factors including race and ethnicity [10]. For instance, Black and Hispanic persons living with dementia are more likely to rely on informal care from family and friends, live in a home with 3 or more household members, and have an adult child in the caregiving network than non-Hispanic White persons living with dementia [14,15]. Despite the importance of caregiver networks, current interventions to support medication management, like providing access to the medical record and using application-based medication trackers, are focused on the person living with dementia–caregiver dyad [8,16]. Networked interventions have the potential to improve adherence and reduce medication administration errors by including the full team of caregivers who help to manage medications [17].

While prior work has highlighted the importance of caregiver networks for providing care, and in communicating between members of caregiver networks when helping with medications [13], the characteristics of the caregiver networks that manage medications for persons living with dementia (the number of people helping with medications, the relationship between the persons living with dementia and the caregivers who help with medications, and the characteristics of those caregivers), and whether there are differences in the characteristics of those who manage medications for persons living with dementia by race and ethnicity, are currently unknown. Hence, the objective of this analysis was to describe the characteristics of caregiver networks that manage medications for persons living with dementia and determine what factors may be associated with helping with medications in caregiver networks.

Methods

Study Design and Setting

This cross-sectional secondary data analysis used data from the 2011 through 2022 National Health and Aging Trends Study (NHATS) “other person” files from rounds 1 through 12. The NHATS is an annual health-related study sampling Medicare beneficiaries in the United States [18]. The NHATS collects detailed information on participants’ physical and cognitive capacity, demographics, and living situations.

Data Source

The NHATS “other person” file is constructed from a roster that is generated from questions in the main NHATS interview and survey, where NHATS respondents were asked about other persons who help with their care. Other persons are cumulative to NHATS participants across rounds, meaning additional other persons are added each round an NHATS participant is interviewed, and each other person is given a unique identifier and resampled with the NHATS participant. While the “other person” file does not specifically identify the demographic features of the NHATS participant, we derived this information by linking the “other person” files with the NHATS participant identifier. NHATS participants and caregivers were analyzed in their most recent year of participation, and data from all waves of NHATS were pooled to increase sample size and estimate the effect of the sampling round on medication management.

Ethical Considerations

This analysis was determined as “not human subjects research” by the Institutional Review Board at the University of Alabama at Birmingham (#300013016). Analytic procedures were determined to be minimal risk to participants and did not require consent or contact with participants, and initial data collection for each NHATS round was approved by the conducting team. Data were deidentified prior to examination and analysis. The reporting of this study adheres to the RECORD (Reporting of Studies Conducted Using Observational Routinely Collected Data) statement (Checklist 1).

Inclusion Criteria

For inclusion as a caregiver in our analysis, caregivers in the NHATS other person file must have been identified as a person who helps with medication management by the NHATS participant or their proxy. The caregiver in the NHATS other person file must also provide care to a patient with dementia, as defined in the NHATS Condition Variable 9, which asks about the presence or absence of a dementia diagnosis.

XSL•FO
RenderX

Assessments

Demographic covariates for caregivers were selected from those collected in the NHATS other person file and included race, ethnicity, gender, relationship to persons living with dementia, education, age, and year of data collection. Race and ethnicity were categorized as White, Black or African American, Hispanic, or other, which included American Indian, Alaska Native, Asian, Native Hawaiian, and Pacific Islander.

Relationship to the persons living with dementia was categorized by self-report as sibling, child, grandchild, spouse or partner, paid helper, or other, which included the sons of step-siblings, boarders and renters, neighbors and nonrelated friends, nonrelated roommates, and ex-spouses or partners. A "paid helper," according to the National Study of Caregiving (NSOC), is someone who was employed to provide care but was not an immediate family member.

We identified caregiver groups who help with medications as those in the other person file who help with the same patient's medications, as determined by an affirmative response to the variable "MEDSHELP," which asks whether the person helps with the patient's medications.

Study Size

We analyzed 15,809 unique caregivers from the "other person" files of NHATS rounds 1 through 12.

Data Analysis

Mixed-effect logistic regression analysis was conducted based on the outcome variable "helping with medications" and independent demographic variables race, relationship to the patient, gender, and education, including a person living with dementia-level random effect to account for multiple caregivers caring for the same person living with dementia. As this was an exploratory rather than confirmatory analysis, variables were not added incrementally. To protect against overfitting and collinearity, we tested for collinearity between independent variables using the variance inflation factor (VIF) and defined problematic collinearity as a VIF greater than 10. We did not impute or otherwise include caregivers who provided missing data. We excluded covariates with problematic missingness, defined as having greater than 20% of missing observations, via listwise deletion to preserve sample size in the regression modeling [19]. Education and age were thus excluded from regression modeling due to missingness, though sensitivity analysis indicated no statistically significant difference between models including and excluding those variables. A predefined α level of .05 was used to determine statistical significance. Odds ratios were converted to predicted probabilities using marginal standardization to aid in interpretability [20]. All analyses were conducted in R version 3.4.1 (The R Foundation for Statistical Computing).

Results

Descriptive Results

Of the caregivers analyzed, 19.3% (3048/15,809) helped manage medications for persons living with dementia. Each person with dementia had at least 1 person who helped manage their medications. Caregivers who helped with medications were 56.9% (1633/2865) non-Hispanic White, 27.8% (799/2865) Black or African American, 9.8% (279/2865) Hispanic, and 4.7% (133/2865) other. Caregivers who helped with medications were 15.0% (445/3033) spouses, 39.1% (1184/3033) children, 4.9% (149/3033) grandchildren, 9.9% (303/3033) other relations, 30.4% (921/3033) paid help, and 1.5% (45/3033) siblings. The mean number of caregivers assisting a person living with dementia with their medications was 2.5, with a range of 1 to 9.

Among non-Hispanic White caregivers who manage medications (n=1633), the mean number of participants assisting with medications was 2.4 (range 1-9). Caregivers who helped with medications were 17.3% (282/1633) spouses, 33.9% (555/1633) children, 2.9% (48/1633) grandchildren, 9.6% (157/1633) others, 35.5% (580/1633) paid helpers, and 0.7% (11/1633) siblings.

Among Black or African American caregivers (n=799), the mean number of caregivers assisting with medications was 2.8 (range 1-9). Caregivers who helped with medications were 8.8% (71/799) spouses, 45.6% (365/799) children, 8.3% (67/799) grandchildren, 12.0% (96/799) others, 22.1% (177/799) paid helpers, and 2.6% (21/799) siblings.

Among Hispanic or Latino caregivers (n=279), the mean number of caregivers assisting with medications was 2.9 (range 1-8). Caregivers who helped with medications were 16.1% (45/279) spouses, 48.7% (136/279) children, 5.3% (15/279) grandchildren, 7.1% (20/279) others, 20.7% (58/279) paid helpers, and 1.7% (5/279) siblings.

Among "other" caregivers (n=133), the mean number of caregivers assisting with medications was 2.4 (range 1-7). Caregivers who helped with medications were 16.5% (22/133) spouses, 31.5% (42/133) children, 5.2% (7/133) grandchildren, 10.5% (14/133) others, 33.1% (44/133) paid help, and 3.0% (4/133) siblings.

Participant demographics can be found in Table 1. See Table 2 for more information on participants by race, ethnicity, and relationship.

XSL•FO
RenderX

Table . Participant characteristics.

	Did not help with medications (n=12,761), n (%)	Helped with medications (n=3048), n (%)	Total (N=15,809), n (%)
Race and ethnicity			
White	7166 (59.0)	1633 (57.4)	8799 (58.7)
Black	3242 (26.7)	799 (28.1)	4041 (27.0)
Hispanic	1183 (9.7)	279 (9.8)	1462 (9.8)
Other	552 (4.5)	133 (4.7)	685 (4.6)
Relationship			
Spouse or partner	284 (2.3)	445 (14.6)	729 (4.7)
Child	5452 (43.2)	1184 (38.9)	6636 (42.4)
Grandchild	1376 (10.9)	149 (4.9)	1525 (9.7)
Other	3789 (30.0)	303 (9.9)	4092 (26.1)
Paid help	809 (6.4)	921 (30.2)	1730 (11.0)
Sibling	906 (7.2)	45 (1.5)	951 (6.1)
Gender			
Male	2160 (48.9)	298 (25.0)	2458 (43.8)
Female	2259 (51.1)	895 (75.0)	3154 (56.2)
Education			
College degree	1180 (25.1)	339 (24.1)	1519 (24.9)
High school or some college	2941 (62.6)	908 (64.5)	3849 (63)
More than a college degree	580 (12.3)	161 (11.4)	741 (12.1)

Table . Number of caregivers helping with medications by race and ethnicity.

	White (n=1633)	Black (n=799)	Hispanic (n=279)	Other (n=133)
Caregivers, mean (range)	2.4 (1-9)	2.8 (1-9)	2.9 (1-8)	2.4 (1-7)
Relationship, n (%)				
Spouse or partner	282 (17.3)	71 (8.9)	45 (16.1)	22 (16.5)
Child	555 (34.0)	365 (45.7)	136 (48.7)	42 (31.6)
Grandchild	48 (2.9)	67 (8.4)	15 (5.4)	7 (5.3)
Other	157 (9.6)	96 (12.0)	20 (7.2)	14 (10.5)
Paid help	580 (35.5)	177 (22.2)	58 (20.8)	44 (33.1)
Sibling	11 (0.7)	21 (2.6)	5 (1.8)	4 (3.0)

Statistical Results

In the univariate statistical model, gender (P<.001) and relationship (P<.001) were associated with helping with medications.

In the fully adjusted regression model, compared to non-Hispanic White race or ethnicity, Black race or ethnicity (predicted probability [PP] 7%, 95% CI 4%-10%; P<.001) and Hispanic race or ethnicity (PP 4%, 95% CI 1%-9%; P=.04) were associated with an increased likelihood of helping with medications among caregivers. Female gender (PP 15%, 95% CI 13%-17%; P<.001) was associated with an increased likelihood of helping with medications. Compared to a spouse or partner, children (PP −51%, 95% CI −56% to −46%; P<.001), grandchildren (PP −60%, 95% CI −66% to −54%; P<.001), paid help (PP −23%, 95% CI −39% to −7%; P=.006), and siblings (PP −63%, 95% CI −69% to −58%; P<.001) were all associated with decreased likelihood of helping with medications. See Table 3 for the fully adjusted regression modeling.

XSL•FO
RenderX

Table . Fully adjusted regression model predicting the probability of helping with medications.

	Predicted probability (95% CI)	*P* value
Female gender	0.15 (0.13 to 0.17)	<.001
Race and ethnicity		
White (reference)		
Black or African American	0.07 (0.04 to 0.1)	<.001
Hispanic	0.04 (0 to 0.09)	.04
Other	−0.01 (−0.07 to 0.05)	.8
Relationship		
Spouse or partner (reference)		
Child	−0.51 (−0.56 to −0.46)	<.001
Grandchild	−0.60 (−0.66 to −0.54)	<.001
Other	−0.62 (−0.67 to −0.57)	<.001
Paid help	−0.23 (−0.39 to −0.07)	.01
Sibling	−0.63 (−0.69 to −0.58)	<.001
Round	0 (0 to 0)	.96

Sensitivity Analysis

We conducted a sensitivity analysis based on reducing the number of distinct relationship categories. In this model, which had relationship categories of spouse or partner, other relative, paid help, and other, we found no improvement from the full regression model when considering Akaike Information Criterion, Bayesian Information Criterion, or overall deviance. See Table 4 for sensitivity analysis.

In subgroup analysis stratified by race and ethnicity, female gender was associated with a higher likelihood of helping with medications, and compared to spouses and partners, children, grandchildren, paid help, and siblings were all associated with a lower likelihood of helping with medications. See Table 5 for subgroup analysis.

Table . Sensitivity results.

Model	AIC[a]	BIC[b]	Deviance
Full model	4415	4493	4391
Sensitive model	4457	4522	4437

[a]AIC: Akaike Information Criterion.

[b]BIC: Bayesian Information Criterion.

XSL•FO
RenderX

Table . Regression models stratified by race or ethnicity predicting the probability of helping with medications.

	Predicted probability (95% CI)	*P* value
White		
Female gender	0.12 (0.09 to 0.15)	<.001
Relationship		
Spouse or partner (reference)		
Child	−0.50 (−0.56 to −0.44)	<.001
Grandchild	−0.57 (−0.65 to −0.5)	<.001
Other	−0.61 (−0.68 to −0.55)	<.001
Paid help	−0.19 (−0.39 to 0)	.05
Sibling	−0.64 (−0.71 to −0.57)	<.001
Round	0 (−0.01 to 0)	.19
Black		
Female gender	0.21 (0.16 to 0.25)	<.001
Relationship		
Spouse or partner (reference)		
Child	−0.48 (−0.59 to −0.36)	<.001
Grandchild	−0.59 (−0.72 to −0.46)	<.001
Other	−0.59 (−0.72 to −0.47)	<.001
Paid help	−0.27 (−0.62 to 0.07)	.12
Sibling	−0.60 (−0.73 to −0.47)	<.001
Round	0 (−0.01 to 0.01)	.59
Hispanic		
Female gender	0.17 (0.11 to 0.24)	<0.001
Relationship		
Spouse or partner (reference)		
Child	−0.65 (−0.78 to −0.53)	<.001
Grandchild	−0.80 (−0.95 to −0.65)	<.001
Other	−0.76 (−0.91 to −0.6)	<.001
Paid help	−0.87 (−0.99 to −0.75)	<.001
Sibling	−0.61 (−0.89 to −0.33)	<.001
Round	0.01 (0 to 0.02)	.09
Other		
Female gender	0.17 (0.07 to 0.27)	<.001
Relationship		
Spouse or partner (reference)		
Child	−0.56 (−0.79 to −0.32)	<.001
Grandchild	−0.56 (−0.84 to −0.27)	<.001
Other	−0.58 (−0.84 to −0.33)	<.001
Paid help	0.30 (0.07 to 0.53)	.01
Sibling	−0.70 (−0.93 to −0.47)	<.001
Round	0.01 (−0.01 to 0.02)	.27

XSL•FO
RenderX

Discussion

Principal Results

In our analysis of 15,809 NHATS "other person" data records from 2011 to 2022, we found that the average person with dementia has 2.5 people that help with their medications, and that Black or African American and Hispanic persons with dementia have more people helping with their medications than White persons with dementia. Many children, grandchildren, and siblings of persons with dementia are helping with medications, in addition to spouses. In the regression modeling, we found that Black race, Hispanic ethnicity, and female gender were associated with an increased likelihood of helping with medications, while relationships other than spouses or partners were associated with a lower likelihood of helping with medications. We did not observe differences between helping with medications and the year of data collection.

Limitations

This study was conducted using a large composite sample from a methodologically robust sample of Medicare beneficiaries. While we included participants from all sampled years (2011 - 2022), we analyzed the sample cross-sectionally at the most recent year of data collection to ensure as large of a sample as possible, though our regression modeling is not generalizable to the broader population as we could not appropriately weigh our combined sample. We included sampling year as a covariate to minimize differences in year on our estimates and found that sampled year was not a statistically significant covariate for medication management, but future work should consider longitudinal trends in medication management by family caregiver networks. Additionally, we chose to exclude education level and age from our regression models due to missingness to preserve sample size on our primary covariates of interest. Prior work has not shown that education or age is associated with medication management among individual caregivers [7], and when our analysis included those variables, we found no statistically significant differences between the 2 models, but future work could consider whether those covariates affect helping with medications in caregiving networks. It is possible that the gender composition of our sample, which was 75% female and 25% male, may overestimate female gender being a statistically significant predictor in our regression modeling. Other work has highlighted the importance of including male dementia caregiver perspectives in research [21]; future targeted work to determine the effect of gender on predictors of network composition should consider oversampling male caregivers to limit this potential effect. Finally, we were unable to include the stage or severity of dementia, or what tasks are included in helping with medications, as this is not collected in NHATS. Future work should consider whether there are differences in caregiver networks that manage medications by stage of dementia, and identifying what tasks of helping with medications are being performed by different components of the network.

Comparison With Prior Literature

Prior work with NHATS data from 2011 to 2015 has found that persons living with dementia have caregiver networks that assist with the tasks of care, which includes medication management as well as other medical-related tasks like attending clinic visits and interacting with providers [22]. They found that these networks range in size from 0 to 2+ caregivers but did not specify a total range. Our finding with data from 2011 to 2022 that caregiver networks included at least one person and, on average, 2.5 caregivers that manage medications indicates that medication management could be an activity that is more complex and requires more caregiver involvement than other caregiving tasks. The differences in network size that we found, while small, highlight the shared nature of caregiving. Not one group had an average of below 2.5 caregivers who manage medications, yet current methods of facilitating medication management are focused on an individual caregiver managing medications. Interventions to support medication management should consider supporting caregiver networks as well as the primary caregiver, as the average person living with dementia has at least 2 caregivers who help with medications. Potential interventions could include applications or integrations that allow for proxy access by caregivers or caregiver profiles that allow for the management of patient medications; these interventions have been considered by health systems, though not applied to medication management [23,24].

Other work on cultural differences in the composition of caregiver networks used data from NHATS from 2015 to 2019 and found differences in the size of caregiver networks by race and ethnicity. They found that Hispanic persons living with dementia had the largest care networks with 6.89 members, while non-Hispanic Black persons living with dementia had 6.69 members and non-Hispanic White persons living with dementia had 5.74 members [10]. Our results indicate that the size of the networks that manage medications are smaller than those that help with other tasks like transportation or cooking food, with Hispanic persons living with dementia having 2.9 members, non-Hispanic Black persons living with dementia having 2.8 members, and non-Hispanic White persons living with dementia having 2.4 members. The relative size of the caregiver networks that manage medications could be due to cultural factors: existing literature has shown differences in family size and cohabitation by race and ethnicity which can affect caregiver network size, and larger networks could have more caregivers helping with medications [14]. Network size could be influenced by living situation: Hispanic persons living with dementia and Black or African American persons living with dementia are less likely to live alone and more likely to live in multigenerational homes than non-Hispanic White persons living with dementia [25,26]. Multigenerational living situations could be affecting the size and composition of the networks of caregivers who help with medications.

Addressing cultural factors is critical in providing culturally specific interventions to support caregivers in medication management. Co-designed medication management materials, where interventions and protocol materials are developed collaboratively between persons living with dementia, caregivers, and researchers, have been identified as an informational gap by prior reviews [27]. Our study suggests that including multiple caregivers may be beneficial to co-designing future interventions, especially in Black and Hispanic communities that have more caregivers on average

XSL•FO
RenderX

and different conceptualizations of the role of caregiving [28,29].

Our finding that spousal caregivers of persons living with dementia were more likely to manage medications when controlling for race and ethnicity and within each race and ethnicity extends prior work which found that spouses were more likely to manage medications for their spouses or partners in a sample that controlled for dementia status [30]. Spouses are more likely to live with the person receiving care and be near the person living with dementia around medication times [31]; however, it is notable that our analysis found that children and grandchildren are the plurality of caregivers who help with medications. While spousal caregivers could be most involved in the administration of medications in the home, children or grandchildren could be instrumental in filling or refilling medication prescriptions. Specific roles and aspects of helping with medications should be the target of additional future studies.

Conclusions

Caregiver network size is an important consideration when designing interventions to support medication management by family caregivers, and this study sheds light on the size and composition of family caregiver networks that manage medications for persons living with dementia. Interventions to support medication administration for persons living with dementia should consider caregiver network size and composition to facilitate persons living with dementia adherence to medication prescriptions.

Acknowledgments

No generative or summative artificial intelligence tools or software were used in this project.

Data Availability

Data are available through the National Health and Aging Trends Study. Analysis is available from the corresponding author on reasonable request.

Authors' Contributions

RWRB contributed to the conceptualization, data curation, formal analysis, writing–original draft, and writing–review and editing. FP contributed to the writing–review and editing. PJB contributed to supervision, writing–original draft, and writing–review and editing. ANL contributed to writing–review and editing. MM contributed to writing–review and editing. JNO contributed to writing–review and editing. RJ contributed to supervision, project administration, and writing–review and editing.

Conflicts of Interest

None declared.

Checklist 1
The RECORD (Reporting of Studies Conducted Using Observational Routinely Collected Data) statement—checklist of items, extended from the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) statement.
[PDF File, 113 KB - i-jmr_v14i1e64499_app1.pdf]

References

1. Page A, Etherton‐Beer C, Seubert LJ, et al. Medication use to manage comorbidities for people with dementia: a systematic review. J Pharm Pract Res 2018 Aug;48(4):356-367. [doi: 10.1002/jppr.1403]
2. Gench M, Sawan MJ, Langford A, Gnjidic D. Tools to evaluate medication management for caregivers of people living with dementia: a systematic review. Health Expect 2021 Oct;24(5):1570-1581. [doi: 10.1111/hex.13318] [Medline: 34289214]
3. Lavretsky H. The role of family caregivers and inappropriate medication use in the community-dwelling older adults with dementia. Aging Health 2012 Oct;8(5):457-460. [doi: 10.2217/ahe.12.51]
4. Look KA, Stone JA. Medication management activities performed by informal caregivers of older adults. Res Social Adm Pharm 2018 May;14(5):418-426. [doi: 10.1016/j.sapharm.2017.05.005] [Medline: 28528023]
5. Shiekh SI, Cadogan SL, Lin LY, Mathur R, Smeeth L, Warren-Gash C. Ethnic differences in dementia risk: a systematic review and meta-analysis. J Alzheimers Dis 2021;80(1):337-355. [doi: 10.3233/JAD-201209] [Medline: 33554910]
6. Cotrell V, Wild K, Bader T. Medication management and adherence among cognitively impaired older adults. J Gerontol Soc Work 2006 Oct 17;47(3-4):31-46. [doi: 10.1300/J083v47n03_03]
7. Noureldin M, Plake KS. Correlates of caregivers' involvement in the management of older adults' medications. Res Social Adm Pharm 2017;13(4):840-848. [doi: 10.1016/j.sapharm.2016.09.009] [Medline: 27916540]
8. Lindauer A, Sexson K, Harvath TA. Medication management for people with dementia. Am J Nurs 2017;117(5):S17-S21. [doi: 10.1097/01.NAJ.0000516389.35634.84]
9. Bailey SC, Oramasionwu CU, Wolf MS. Rethinking adherence: a health literacy-informed model of medication self-management. J Health Commun 2013;18(Suppl 1):20-30. [doi: 10.1080/10810730.2013.825672] [Medline: 24093342]

XSL•FO
RenderX

10. Wyman MF, Vega IE, Cabrera LY, Amini R, Kim K, Tarraf W. Influence of care network characteristics on physician visit experiences for Black, White, and Hispanic older adults with dementia. J Gerontol B Psychol Sci Soc Sci 2023 Mar 13;78(Suppl 1):S59-S70. [doi: 10.1093/geronb/gbac150] [Medline: 36205710]
11. Friedman EM, Kennedy DP. Typologies of dementia caregiver support networks: a pilot study. Gerontologist 2021 Nov 15;61(8):1221-1230. [doi: 10.1093/geront/gnab013] [Medline: 33585929]
12. Xu J, Liu PJ, Beach S. Multiple caregivers, many minds: family discord and caregiver outcomes. Gerontologist 2021 Jul 13;61(5):661-669. [doi: 10.1093/geront/gnaa067] [Medline: 32539097]
13. Managing multiple caregivers: 7 tips to stay organized. Division of Pharmacy Professional Development, University of Wisconsin–Madison. 2018. URL: https://ce.pharmacy.wisc.edu/blog/managing-multiple-caregivers-7-tips-to-stay-organized/ [accessed 2025-04-17]
14. Parker LJ, Fabius C. Who's helping whom? Examination of care arrangements for racially and ethnically diverse people living with dementia in the community. J Appl Gerontol 2022 Dec;41(12):2589-2593. [doi: 10.1177/07334648221120247] [Medline: 35960528]
15. Llanque SM, Enriquez M. Interventions for Hispanic caregivers of patients with dementia: a review of the literature. Am J Alzheimers Dis Other Demen 2012 Feb;27(1):23-32. [doi: 10.1177/1533317512439794] [Medline: 22467412]
16. Bratches RWR, Wall JA, Puga F, et al. Patient portal use among family caregivers of individuals with dementia and cancer: regression analysis from the national study of caregiving. JMIR Aging 2023 Dec 20;6(1):e44166. [doi: 10.2196/44166] [Medline: 38235767]
17. Liang J, Aranda MP, Jang Y, Wilber K, Chi I, Wu S. The effect of support from secondary caregiver network on primary caregiver burden: do men and women, Blacks and Whites differ? J Gerontol B Psychol Sci Soc Sci 2022 Oct 6;77(10):1947-1958. [doi: 10.1093/geronb/gbac067] [Medline: 35511820]
18. Kasper JD, Freedman VA. Findings from the 1st round of the National Health and Aging Trends Study (NHATS): introduction to a special issue. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69(Suppl 1):S1-S7. [doi: 10.1093/geronb/gbu125] [Medline: 25342818]
19. Kang H. The prevention and handling of the missing data. Korean J Anesthesiol 2013 May;64(5):402-406. [doi: 10.4097/kjae.2013.64.5.402] [Medline: 23741561]
20. Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. Int J Epidemiol 2014 Jun;43(3):962-970. [doi: 10.1093/ije/dyu029]
21. Bueno MV. A duty to care: male perspectives on the caregiver role for persons with Alzheimer's or dementia. J Fam Nurs 2024 Jan 11. [doi: 10.1177/10748407231222291] [Medline: 38205720]
22. Spillman BC, Freedman VA, Kasper JD, Wolff JL. Change over time in caregiving networks for older adults with and without dementia. J Gerontol B Psychol Sci Soc Sci 2020 Aug 13;75(7):1563-1572. [doi: 10.1093/geronb/gbz065] [Medline: 31102533]
23. Dickman Portz J, Powers JD, Casillas A, et al. Characteristics of patients and proxy caregivers using patient portals in the setting of serious illness and end of life. J Palliat Med 2021 Nov 1;24(11):1697-1704. [doi: 10.1089/jpm.2020.0667]
24. Latulipe C, Quandt SA, Melius KA, et al. Insights into older adult patient concerns around the caregiver proxy portal use: qualitative interview study. J Med Internet Res 2018 Nov 2;20(11):e10524. [doi: 10.2196/10524]
25. Cantu PA, Kim J, López-Ortega M, Rote S, Mejia-Arango S, Angel JL. Living arrangements and dementia among the oldest old: a comparison of Mexicans and Mexican Americans. Innov Aging 2022;6(3):igac014. [doi: 10.1093/geroni/igac014] [Medline: 35663277]
26. Samson ZB, Parker M, Dye C, Hepburn K. Experiences and learning needs of African American family dementia caregivers. Am J Alzheimers Dis Other Demen 2016 Sep;31(6):492-501. [doi: 10.1177/1533317516628518] [Medline: 26953236]
27. Clough AJ, Gnjidic D, Cross AJ, et al. Medication management information priorities of people living with dementia and their carers: a scoping review. Age Ageing 2024 Sep 1;53(9):afae200. [doi: 10.1093/ageing/afae200] [Medline: 39287004]
28. Pan Y, Chen R, Yang D. The relationship between filial piety and caregiver burden among adult children: a systematic review and meta-analysis. Geriatr Nurs (Lond) 2022 Jan;43:113-123. [doi: 10.1016/j.gerinurse.2021.10.024]
29. Allen AP, Curran EA, Duggan Á, et al. A systematic review of the psychobiological burden of informal caregiving for patients with dementia: focus on cognitive and biological markers of chronic stress. Neurosci Biobehav Rev 2017 Feb;73:123-164. [doi: 10.1016/j.neubiorev.2016.12.006] [Medline: 27986469]
30. Polenick CA, Stanz SD, Leggett AN, Maust DT, Hodgson NA, Kales HC. Stressors and resources related to medication management: associations with spousal caregivers' role overload. Gerontologist 2020 Jan 24;60(1):165-173. [doi: 10.1093/geront/gny130] [Medline: 30358854]
31. Reinhard S. Family caregivers providing complex care. Innov Aging 2019 Nov;3(Suppl 1):S747-S748. [doi: 10.1093/geroni/igz038.2740]

Abbreviations

NHATS: National Health and Aging Trends Study
PP: predicted probability

XSL•FO
RenderX

RECORD: Reporting of Studies Conducted Using Observational Routinely Collected Data
VIF: variance inflation factor

Edited by TDA Cardoso; submitted 18.07.24; peer-reviewed by M Lopez-Anuarbe, MJ Sawan; revised version received 13.03.25; accepted 14.03.25; published 22.05.25.

Please cite as:
Bratches RWR, Puga F, Barr PJ, Leggett AN, Masel M, Odom JN, Jablonski R
Size and Composition of Caregiver Networks Who Manage Medications for Persons Living With Dementia: Cross-Sectional Analysis of the 2011-2022 National Health and Aging Trends Study
Interact J Med Res 2025;14:e64499
URL: https://www.i-jmr.org/2025/1/e64499
doi:10.2196/64499

© Reed WR Bratches, Frank Puga, Paul J Barr, Amanda N Leggett, Meredith Masel, James Nicholas Odom, Rita Jablonski. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.5.2025. 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 Prioritization of Behavior Change Techniques for Long-Term Stroke Rehabilitation: Delphi Survey Study

Agata Ewa Wróbel[1], PhD; Philip Cash[2], PhD; Anja Maier[1,3], PhD; John Paulin Hansen[1], PhD

[1]Department of Technology, Management and Economics, DTU Technical University of Denmark, Lyngby, Denmark

[2]Design School, Northumbria University, Newcastle, United Kingdom

[3]Department of Design, Manufacturing and Engineering Management, University of Strathclyde, Glasgow, United Kingdom

Corresponding Author:
Anja Maier, PhD
Department of Technology, Management and Economics, DTU Technical University of Denmark, Lyngby, Denmark

Abstract

Background: Stroke results in both physical disability and psychological distress. The impact can be minimized through rehabilitation, but it is a long-term process, making it difficult for patients to adhere to treatment. Thus, a better understanding of long-term behavior change interventions for patients with stroke is needed as well as how such interventions can support not only rehabilitation of motoric functions but also mental well-being.

Objective: The aim of this study is to understand both the most important behavior change technique (BCT) clusters for long-term stroke rehabilitation in general as well as which are most relevant for each aspect of stroke rehabilitation: behavioral, cognitive, and emotional.

Methods: We applied the 16 BCT clusters. The study used a 2-round Delphi survey, as reliable consensus was obtained among a group of 12 international experts. Experts represented three main backgrounds involved in behavioral intervention in the health context: (1) specialists in behavioral science (n=4), (2) behavioral designers (n=4), and (3) expert health care professionals (n=4). Experts were brought together in this way for the first time. In the first round, web-based questionnaires were used to collect data from the experts. This was followed by a personalized second round. Consensus was determined by statistically aggregating the responses and evaluating IQR and percentage consensus. BCT clusters reaching consensus (IQR ≤1 and percentage ≥50%) were then ranked.

Results: In total, 12 of 16 BCT clusters reached consensus for general importance in stroke rehabilitation, with 11, 9, and 6 BCT clusters achieving consensus for, respectively, the behavioral, cognitive, and emotional aspects of rehabilitation. The overall most relevant BCT clusters were repetition and substitution, social support, feedback and monitoring, and self-belief, with similar outcomes for behavioral and cognitive rehabilitation. For emotional rehabilitation, social support and identity were emphasized. The least relevant BCT clusters were natural consequences, covert learning, and comparison of behavior.

Conclusions: This expert panel study using a 2-round Delphi survey ranked the importance of BCT clusters for long-term stroke rehabilitation. The process yielded a number of novel insights highlighting differences in importance between general rehabilitation and that specifically focused on the behavioral, cognitive, and emotional aspects of stroke recovery. This provides a first but important step toward unlocking the prioritization of BCT clusters for long-term intervention contexts such as stroke rehabilitation and enables effective intervention mapping addressing long-term behavior change and treatment adherence.

(Interact J Med Res 2025;14:e59172) doi:10.2196/59172

KEYWORDS

behavior change; behavior change techniques; BCT; neuroscience; neurology; stroke; rehabilitation; adherence; Delphi method; Delphi; intervention design; intervention mapping

Introduction

Background

The World Health Organization defines stroke as "(...) rapidly developed clinical signs of focal or global disturbance of cerebral function, lasting more than 24 hours or until death, with no apparent non-vascular cause" [1]. It is a serious health problem and one of the most common causes of death and acquired disability among adults [2], having "(...) the greatest disabling impact of any chronic disease" [3]. Further, only a small proportion of patients die in the acute stroke phase, leaving the majority with moderate to severe disability that can be mitigated through early and sustained rehabilitation interventions [4]. It is estimated that the number of people living with stroke will increase by 27% in the European Union alone between 2017 and 2047 [5]. As such, stroke has a considerable socioeconomic impact, affecting not only patients themselves

XSL•FO
RenderX

but also their families, caretakers, medical professionals, hospitals, policy makers, and the whole health care system in the long term [6]. This places a critical focus on effective poststroke treatment and rehabilitation [7] not least for making health care interventions usable, safe, and effective.

Patients with stroke experience both physical disability and psychological changes, including emotional, behavioral, and cognitive impairments [3]. Issues, including depression, apathy, anxiety, posttraumatic stress, and anger, among others [3], are prevalent in patients with stroke. These pose significant barriers to rehabilitation by negatively impacting long-term adherence to exercises, hindering treatment processes, and decreasing the chance of patient recovery [8,9]. This places a priority on positively shaping patient behavior change in the long term. Yet, Teasell et al [10] highlight that over two-thirds of interventions in this context focus on improving motor functions, while less than 6% dealt with psychosocial issues. As such, there is a need to highlight the relevance of behavior change techniques (BCTs) in addressing psychosocial issues. However, general discussions of BCTs typically focus on interventions framed in the short term [11]. For example, Liu et al [12] highlight how a single reward can be effective in fostering short-term sign-up behavior, but additional strategies and research are needed to understand how best to maintain engagement. This uncovers a gap between current stroke rehabilitation practice and the potential of BCTs, as it leaves critical questions as to which BCTs should be prioritized when supporting the long-term treatment adherence required. Further, this reinforces the relevance and urgency of addressing long-term behavior change and treatment adherence specifically for stroke rehabilitation.

Together, these challenges reveal a critical disconnect between current stroke rehabilitation interventions and more general guidance on BCT use [13]. Thus, this work focuses on resolving this disconnect by examining how BCTs should be prioritized for stroke rehabilitation with a multifaceted approach that distinguishes between motor rehabilitation ("general rehabilitation" in our terms) and psychosocial rehabilitation (addressed separately as "behavioral," "cognitive," and "emotional" aspects). As such, the aim of this study is to understand both the most important BCT clusters for long-term stroke rehabilitation in general as well as which are most relevant for each aspect of stroke rehabilitation: behavioral, cognitive, and emotional.

Patient Adherence to Treatment

Poor adherence to treatment is a common challenge across health care domains [8]. In physiotherapy, up to 65% (with some sources claiming even 70%) of patients are nonadherent or only partially adherent to home treatment and exercises [8,14]. This high level of nonadherence (noncompliance) has an adverse effect on the cost and effectiveness of rehabilitation [8,15]. Several studies to date have explored the reasons for patient nonadherence to home treatment and exercises [8,14,16,17]. In their review, Jack et al [8] found strong support for the following barriers to treatment adherence: "low levels of physical activity at baseline or in previous weeks, low in-treatment adherence with exercise, low self-efficacy, depression, anxiety, helplessness, poor social support or activity, greater perceived number of barriers to exercise, and increased pain levels during exercise." However, while the importance of identifying and acknowledging the above during patient assessment has been emphasized, few authors offer concrete solutions to address adherence.

Two notable exceptions to this are Campbell et al [15] and Bonnechre et al [16]. Campbell et al [15] note the importance of physiotherapist's presence and contact with the patient as well as including patients as partners in planning therapy; and Bonnechre et al [16] stress the positive effect of patient supervision and suggest that, for example, telerehabilitation or other monitoring systems could be beneficial. However, when compared to the scope of relevant BCTs [18], these examples primarily serve to highlight the untapped potential of behavior changing interventions in support of long-term stroke rehabilitation. First, they are not specific to stroke rehabilitation; hence, they do not address some key issues that patients with stroke deal with compared to other conditions, such as simultaneous impairment of both motoric and cognitive functions. This makes the rehabilitation process more complex and increases the risk of developing other conditions such as depression and anxiety. Second, many approaches to improve patient adherence lack individualization, which is crucial as every patient is "(...) structurally, chemically, and emotionally different" [19]. This is especially true in the case of stroke, which can occur at different ages (despite being more prevalent in older age groups), be located in different regions of the brain with different functional consequences, and yield a range of effects, from mild to very severe [5]. Therefore, potential interventions must be considered in terms of long-term feasibility, as well as suitability for home treatment and rehabilitation, with minimum assistance of third persons. Each of these challenges points to potential mechanisms found in BCTs.

BCTs in Health Care

BCTs used as the basis for health care interventions such as education, training, or facilitating behavior through design, have been successfully applied to drive changes in health care professionals' practice [20,21], patient treatment and rehabilitation [22], disease prevention [23], or self-management [24]. At the same time, while the topic has been studied extensively by both behavioral psychologists [18] and designers [25,26], the differentiation between short- and long-term interventions has been little discussed. This poses a notable issue for rehabilitation, as it is widely acknowledged that sustaining long-term change introduces specific behavioral and design challenges [11,27,28].

A general approach to designing health care interventions suggests the following steps: "(...) identifying barriers, selecting intervention components, using theory, and engaging end-users" [21], linking development to feasibility, evaluation, and implementation [29,30]. In rehabilitation specifically, particular attention is given to the rehabilitation provider, who acts as a facilitator and enabler in driving health behavior change [31]. Nieuwenhuijsen et al [31] also highlight the need to understand a person's context and environmental factors, as these can

XSL•FO
RenderX

become facilitators or barriers to change. While these considerations are essential, guidance on how to approach long-term change or prioritize BCTs in this context is missing. This is particularly relevant in the case of stroke. This is because after the first few months of intense treatment, rehabilitation can take several months or even years in a home setting. Further, due to the complexity and scope of possible BCTs, prioritization efforts are critical to their understanding and application [18,32]. Here, one of the most widely accepted taxonomies of BCTs is provided by Michie et al [18]. Within the taxonomy of Michie et al [18], there are 16 clusters, with 93 individual BCTs, presenting a huge scope for intervention. Therefore, there is an urgent need to better understand which BCT clusters are most relevant to interventions for long-term contexts such as stroke rehabilitation.

Research Question

Bringing the literature together, there is a need to understanding both the most important BCT clusters for long-term stroke rehabilitation in general as well as which are most relevant for each aspect of stroke rehabilitation: behavioral, cognitive, and emotional, that is, supporting not only motoric and cognitive function but also mental well-being. Hence, we address one main research question (RQ): Which BCT clusters are the most relevant for long-term stroke rehabilitation in a home setting in relation to treating its behavioral, cognitive, and emotional aspects?

Methods

Study Design

To address the RQ, we used a Delphi method, which has proven to be essential in ranking importance of BCT clusters in different contexts when large-scale data on intervention outcomes are notably rare [18,32-35].

The Delphi method was administered by email and using a link to a web-based survey tool (Qualtrics) [36]. We followed the procedure set out by Rowe et al [37] and obtained consensus after 2 rounds. The 2-round Delphi survey was used to rank the least and most important BCT clusters across the parameters outlined in the RQ. The 2 rounds were executed in 2021 between June and October. In the first round, the 16 BCT clusters defined in the behavior change technique taxonomy (BCTT; version 1) were used as the basis for the survey structure, and questions were evaluated by 12 international experts [18,33]. BCTT (version 1) was used for two main reasons: (1) it provides a basis for direct comparison with an array of prior ranking studies and hence helps to contextualize the results of this work and (2) it is widely applied in practice, making it accessible to all involved experts. In the second round, the results were iterated by 9 of the original 12 experts to reach a consensus on the ranking of the BCT clusters. The 16 BCT clusters are summarized in Textbox 1 as they were presented to the participants.

XSL•FO
RenderX

Textbox 1. Summary of the 16 behavior change technique (BCT) clusters in behavior change technique taxonomy (version 1) as presented to participants.

1. Goals and planning comprises BCTs that revolve around goal-setting and revision, strengthening commitment to those goals, problem-solving (by analyzing and removing factors influencing certain behavior), and action planning.
2. Feedback and monitoring comprises BCTs that revolve around monitoring (by others) or self-monitoring (yourself) the behaviors or behavior outcomes, which can—but does not have to—be combined with feedback on behavior, its outcomes, or body state (biofeedback).
3. Social support comprises BCTs that revolve around providing different types of social support to assist the person in the process of behavior change.
4. Shaping knowledge comprises BCTs that revolve around providing information on the antecedents of the behavior (including suggesting alternative explanations) and instructions on how to perform the behavior.
5. Natural consequences comprises BCTs that revolve around providing information (and occasionally also monitoring) on the health, social, environmental, and emotional consequences of the person's behavior.
6. Comparison of behavior comprises BCTs that revolve around comparing person's behavior to a correct behavior or others' behavior as well as informing about what other people think about the behavior.
7. Associations comprises BCTs that revolve around providing prompts or cues as a stimulus for the behavior and gradually removing them when the goal is achieved. The cues can also be related to signaling the reward (or lack of thereof) or removing a stimulus that causes adverse effects.
8. Repetition and substitution comprises BCTs that revolve around practicing or rehearsing behaviors (sometimes in an excessive manner—overcorrection) to increase habit and skill, as well as behavior substitution with a wanted or neutral behavior, and habit formation or reversal.
9. Comparison of outcomes comprises BCTs that revolve around providing objective, credible source of information in favor or against the behavior, as well as advising comparison of pros and cons of the behavior or possible outcomes of doing (vs not doing) a given behavior.
10. Reward and threat comprises BCTs that revolve around providing a material or social reward or incentive to encourage the desired behavior. The subtle difference between a reward and incentive refers to the fact that the former is actually performed at a given time (eg, arranging for the person to receive money that would have been spent on cigarettes if and only if the smoker has not smoked for 1 month would be a material reward, while informing that the payment will be made in the future for every month a person did not smoke during pregnancy would be a material incentive). Similarly, a reward or incentive can relate to the behavior itself (eg, sticking to a diet every day for a month) or the outcome of the behavior (eg, losing 5 kg in a month). Finally, the rewards or incentives can be provided externally or by the person themselves (eg, plan to reward yourself with new clothes if you adhered to a healthy diet).
11. Regulation comprises BCTs that revolve around providing different ways to regulate and facilitate one's behavior. That could include encouraging the use of drugs that facilitate behavior change, advising on ways of reducing negative emotions or minimizing mental resources to support behavior change, or giving paradoxical instructions to engage in the unwanted behavior with a point of reducing motivation to do it (eg, asking a smoker to smoke twice as many cigarettes a day than they usually do or tell the person to stay awake as long as possible in order to reduce insomnia).
12. Antecedents comprises BCTs that revolve around changing one's physical or social environment to facilitate the performance of wanted behaviors or create barriers for unwanted behaviors. This can also include adding objects to the environment, advising on how to avoid or reduce exposure to specific social or physical cues that trigger unwanted behavior, or how to distract attention with other activities. It could also mean introducing body changes to facilitate behavior change (eg, strength training).
13. Identity comprises BCTs that revolve around attention to self-image related to the given behavior. It could be informing the person that their behavior may be an example to others, suggesting deliberate change of perspective on behavior (framing or reframing) to change emotions associated with the behavior or drawing attention to differences between the behavior and person's self-image. It could also include asking the person to construct and articulate their new identity (eg, as an "ex-smoker").
14. Scheduled consequences comprises BCTs that revolve around the planned consequences of the behavior, which could be both positive (eg, rewarding completion [or even approximation] of the behavior) or negative (eg, punishing the unwanted behavior or removing the reward).
15. Self-belief comprises BCTs that revolve around the idea of reinforcing one's belief that they can and will successfully perform the wanted behavior and arguing against self-doubt and insecurity.
16. Covert learning comprises BCTs that revolve around imagining performing the given behavior and its consequences or observing the consequences of this behavior for others.

Delphi Panel

For the Delphi panel, we convened 12 international experts on behavioral intervention and stroke rehabilitation based on the scientific, medical, and industrial networks of the authors, following other recent Delphi studies in this area [33,38]. Here, expert was defined as "someone with knowledge and experience on a particular subject matter" [39]. The experts were identified through the academic and professional networks of the authors (focusing on those in the fields of design and clinical practice), with eligibility criteria as follows. All experts had at least a master degree as well as multiple years of experience in practice. Further, they all had expertise in long-term behavioral design, development of interventions in this context, and familiarity with the BCTT, that is, they were able to distinguish the BCTT clusters and their implications. While these criteria led to a relatively small sample, there are several reasons for this. First, the population deeply familiar with both the full BCTT and stroke education and rehabilitation is very limited, with several of our contacts rejecting our invitation on this basis despite

XSL•FO
RenderX

being relevant subject experts. This is a familiar issue for studies of BCT use; thus, our experience and sample align with an array of recent BCT-related Delphi studies [33,34]. Second, we specifically include experts focused on the development of interventions in the emerging area of behavioral design, where there are even less experienced practitioners [26]; thus, in order to maintain a balanced sample, each group was limited based on this smallest population. This subgroup was specifically included due to its growing impact on the development and implementation of health interventions [20,34]. Thus, of the 26 experts approached, we accepted a sample of 12 who agreed to participate in the study.

Overall, 12 completed the first round and 9 (75%) the second round, exceeding the minimum 70% threshold for Delphi round response [40]. Experts represented three main backgrounds involved in behavioral intervention in the health context: (1) specialists in behavioral science, design for behavior change, or psychology, who research BCTs (n=4); (2) behavioral designers working in industry actively developing health-related behavioral interventions (n=4); and (3) expert health care professionals working with patients with stroke daily (n=4). The participants were residents of Denmark (8), Germany (2), and Portugal (2). Given the heterogeneous group, this panel is in line with recommendations for a number of participants [41,42] as well as comparable studies in this area [33,38].

Ethical Considerations

The ethics approval processes at the host research institution and national guidelines were followed. No further ethics approval was needed with reference to the Danish Research Ethics Committees that state on their website that "(...) health science questionnaire surveys and interview studies that do not involve human biological material (section 14(2) of the Danish Act on Committees) are not mandatory to report" [43].

Procedure

Delphi Study First Round

The first round of the Delphi study used a web-based questionnaire distributed via Qualtrics software. The questionnaire consisted of 3 parts. In the first part, the participants were asked to provide demographic information, such as age, gender, education, professional background, type of employment, and country of residence (which in the health care context provides more information than nationality, as clinical practice and rehabilitation process differs depending on location). The second part of the questionnaire consisted of 3 open questions, where the participants could elaborate on the context of their current job or research, their understanding of "behavior change" in the context of their work, as well as what they typically consider a "long-term" change. The third and main part of the questionnaire was divided into sections corresponding to the 16 BCT clusters [18]. Each section started with a description of a given cluster (Textbox 1) as well as the specific BCTs in the cluster and examples of practical interventions using it. The description was followed by 2 questions focused specifically on how the BCTs from this cluster support stroke rehabilitation and treatment of psychological issues after stroke, including emotional, behavioral, and cognitive recovery. The questions used a standard 5-point Likert scale (rated from 1=not at all important to 5=extremely important) and are summarized below:

1. In your opinion, how important is [*BCT cluster*] in the context of rehabilitation of patients with stroke?
2. In your opinion, how important is [*BCT cluster*] for the treatment of the following psychological issues after stroke?
   a. Behavioral (eg social judgment and personality changes)
   b. Cognitive (eg attention, memory, or language impairment)
   c. Emotional (eg, depression, anxiety, posttraumatic stress, and anger)

At the end of the questionnaire, there was an opportunity for the experts to share their impressions on the topic and any comments that arose during the study.

Delphi Study Second Round

After the results from the first round were analyzed, a second questionnaire was designed for each specific participant, delivered via a personalized email. This also provided feedback on the responses to the first round for participants to reconsider their prior answers. The purpose of this round was to determine consensus on only those BCT clusters where no consensus was reached in the first round. This limited the scope of the second round in order to ease participation and followed similar approaches in prior work on BCT ranking [34]. In each case, participants had the opportunity to change their prior answers in light of the first round responses or to provide additional explanations of their reasoning. The explanation of the questions and BCT clusters was the same as in the first round.

Data Analysis: Consensus Evaluation

After both rounds of the Delphi survey were complete, the final level of consensus and ranking of the BCT clusters was determined. While the means for establishing consensus and its level in Delphi studies are debated, we apply two widely acknowledged criteria: (1) percentage consensus [44] and (2) median and IQR [33,45]. In terms of overall percentage consensus, a minimum of ≥50% is considered acceptable and provides context for the IQR evaluation [44]. When using Likert scales in Delphi studies, agreement with respect to a median score is widely evaluated with respect to IQR. Specifically, for a 5-point Likert scale, an IQR of ≤1 is suggested as indicative of a high degree of consensus [44,45]. Thus, in this study, we considered consensus to be reached when both percentage consensus was ≥50% and IQR was ≤1, that is, the majority of answers fell within 1 point on the 5-point Likert scale. The percentage consensus, median, and IQR results are summarized for all BCT clusters in Table 1 (for survey question 1) and Table 2 (for survey question 2a, b, and c).

XSL•FO
RenderX

Table . The percentage consensus and median (IQR) for all behavior change technique (BCT) clusters related to question 1 after round 2.

BCT cluster	Consensus	Median (IQR)
1. Goals and planning	67	4.0 (4.0-4.3)
2. Feedback and monitoring	42	4.0 (4.0-5.0)
3. Social support	50	4.5 (4.0-5.0)
4. Shaping knowledge	58	4.0 (3.8-4.0)
5. Natural consequences	50	3.0 (2.0-3.0)
6. Comparison of behavior	50	2.0 (2.0-3.0)
7. Associations	42	4.0 (3.0-4.0)
8. Repetition and substitution	50	4.5 (4.0-5.0)
9. Comparison of outcomes	67	4.0 (2.8-4.0)
10. Reward and threat	33	3.0 (3.0-4.0)
11. Regulation	50	3.0 (3.0-4.0)
12. Antecedents	50	4.0 (3.0-4.0)
13. Identity	42	3.5 (2.8-4.0)
14. Scheduled consequences	42	2.5 (2.0-4.0)
15. Self-belief	50	4.0 (4.0-5.0)
16. Covert learning	58	3.0 (2.8-3.0)

Table . The percentage consensus and median (IQR) for all behavior change technique (BCT) clusters related to question 2a, 2b, and 2c after round 2 (clusters are numbered for brevity).

BCT	Question 2a		Question 2b		Question 2c	
	Consensus	Median (IQR)	Consensus	Median (IQR)	Consensus	Median (IQR)
1	50	3.5 (3.0-4.0)	33	4.0 (3.0-5.0)	50	4.0 (3.0-4.0)
2	50	4.0 (3.8-4.3)	42	4.0 (3.8-5.0)	50	4.0 (3.0-4.0)
3	58	4.0 (3.5-4.0)	58	4.0 (3.0-4.0)	67	5.0 (4.0-5.0)
4	58	4.0 (3.0-4.0)	42	4.0 (3.0-4.0)	50	4.0 (2.8-4.0)
5	42	3.0 (2.0-3.3)	42	2.0 (2.0-3.0)	67	2.0 (2.0-4.0)
6	42	3.0 (2.0-4.0)	42	3.0 (1.8-3.0)	25	2.5 (1.8-3.3)
7	50	4.0 (3.0-4.0)	50	4.0 (3.0-4.0)	42	3.0 (3.0-4.0)
8	67	4.0 (4.0-4.3)	50	4.0 (4.0-5.0)	33	3.0 (2.8-4.0)
9	67	4.0 (2.0-4.0)	42	3.0 (2.0-3.3)	42	4.0 (2.8-4.0)
10	50	3.0 (2.8-3.3)	50	3.0 (2.0-3.0)	33	3.0 (2.0-4.0)
11	50	3.5 (3.0-4.0)	50	3.0 (3.0-4.0)	50	3.5 (2.8-4.0)
12	58	4.0 (3.0-4.0)	50	4.0 (2.8-4.0)	33	3.0 (2.0-4.0)
13	42	3.5 (2.8-4.0)	42	3.0 (2.8-4.0)	50	4.0 (3.8-4.3)
14	33	3.0 (2.0-4.0)	33	2.5 (1.0-3.3)	33	2.5 (2.0-4.0)
15	50	4.0 (3.8-4.3)	67	4.0 (4.0-4.0)	42	4.0 (3.8-5.0)
16	50	3.0 (2.0-3.0)	50	2.0 (2.0-3.0)	58	3.0 (2.0-3.0)

Ranking

Once the level of consensus was determined, the mean and SD were calculated for all BCT clusters that had reached consensus. The mean was then used to rank the importance of each BCT cluster with respect to the parameters in the RQ. This follows prior work evaluating BCT importance [33].

XSL•FO
RenderX

Results

BCT Clusters for Long-Term Stroke Rehabilitation

In answer to which BCT clusters are the most relevant for long-term stroke rehabilitation in a home setting in general, consensus was reached for 12 of 16 BCT clusters after the second round Delphi survey. No consensus was reached for the comparison of outcomes, reward and threat, identity, and scheduled consequences (IQRs>1). The importance of the remaining 12 clusters is summarized in Table 3 based on their means, and with the 3 most and 3 least relevant highlighted in italics format. The most relevant BCT clusters were repetition and substitution (mean 4.50, SD 0.52), social support (mean 4.42, SD 0.67), feedback and monitoring (mean 4.25, SD 0.75), and self-belief (mean 4.25, SD 0.87), while the least relevant were natural consequences (mean 3.00, SD 1.04), covert learning (mean 2.83, SD 0.83), and comparison of behavior (mean 2.50, SD 0.90).

Table . Consensus importance ranking of behavior change technique (BCT) clusters for overall stroke rehabilitation.

Rank	BCT cluster	Mean (SD)
1	8. Repetition and substitution	4.50 (0.52)[a]
2	3. Social support	4.42 (0.67)[a]
3	2. Feedback and monitoring	4.25 (0.75)[a]
3	15. Self-belief	4.25 (0.87)[a]
5	1. Goals and planning	4.17 (0.58)
6	4. Shaping knowledge	3.92 (0.67)
7	12. Antecedents	3.75 (0.87)
8	7. Associations	3.58 (1.00)
9	11. Regulation	3.42 (0.79)
10	5. Natural consequences	3.00 (1.04)[b]
11	16. Covert learning	2.83 (0.83)[b]
12	6. Comparison of behavior	2.50 (0.90)[b]

[a]The 3 most relevant.

[b]The 3 least relevant.

BCT Clusters for Behavioral, Cognitive, and Emotional Aspects of Stroke Rehabilitation

In answer to our RQ regarding which BCT clusters are the most relevant for treating behavioral, cognitive, and emotional aspects, different levels of consensus were reached across the 3 aspects evaluated. For behavioral rehabilitation, 11 BCT clusters reached consensus, while for cognitive, this number was 9 and for emotional 6. The importance of the remaining clusters in each aspect is summarized in Tables 4-6 (behavioral, cognitive, and emotional, respectively). Again, the 3 most and 3 least relevant for each aspect are highlighted in italics format. For both behavioral and cognitive rehabilitation, repetition and substitution (mean 4.17, SD 0.58 and mean 4.33, SD 0.65, respectively) as well as self-belief (mean 4.00, SD 0.74 and mean 3.83, SD 0.94, respectively) were ranked as most relevant, while covert learning was consistently ranked as one of the least relevant BCT clusters across all 3 aspects. The consensus importance results for the 3 aspects are summarized in Figure 1.

XSL•FO
RenderX

Table . Consensus importance ranking of behavior change technique (BCT) clusters for behavioral rehabilitation.

Rank	BCT cluster	Mean (SD)
1	8. Repetition and substitution	4.17 (0.58)[a]
2	15. Self-belief	4.00 (0.74)[a]
3	2. Feedback and monitoring	3.92 (0.90)[a]
4	4. Shaping knowledge	3.75 (0.62)
5	3. Social support	3.67 (1.07)
5	12. Antecedents	3.67 (0.78)
7	1. Goals and planning	3.58 (0.67)
8	7. Associations	3.33 (1.23)[b]
8	11. Regulation	3.33 (0.78)[b]
10	10. Reward and threat	3.00 (0.74)[b]
11	16. Covert learning	2.83 (0.72)[b]

[a]The 3 most relevant.

[b]The 3 least relevant.

Table . Consensus importance ranking of behavior change technique (BCT) clusters for cognitive rehabilitation.

Rank	BCT cluster	Mean (SD)
1	8. Repetition and substitution	4.33 (0.65)[a]
2	15. Self-belief	3.83 (0.94)[a]
3	4. Shaping knowledge	3.67 (0.89)[a]
4	7. Associations	3.58 (1.16)
5	3. Social support	3.42 (0.79)
6	11. Regulation	3.17 (1.03)
7	10. Reward and threat	2.75 (0.87)[b]
8	16. Covert learning	2.50 (0.90)[b]
9	5. Natural consequences	2.33 (0.89)[b]

[a]The 3 most relevant.

[b]The 3 least relevant.

Table . Consensus importance ranking of behavior change technique (BCT) clusters for emotional rehabilitation.

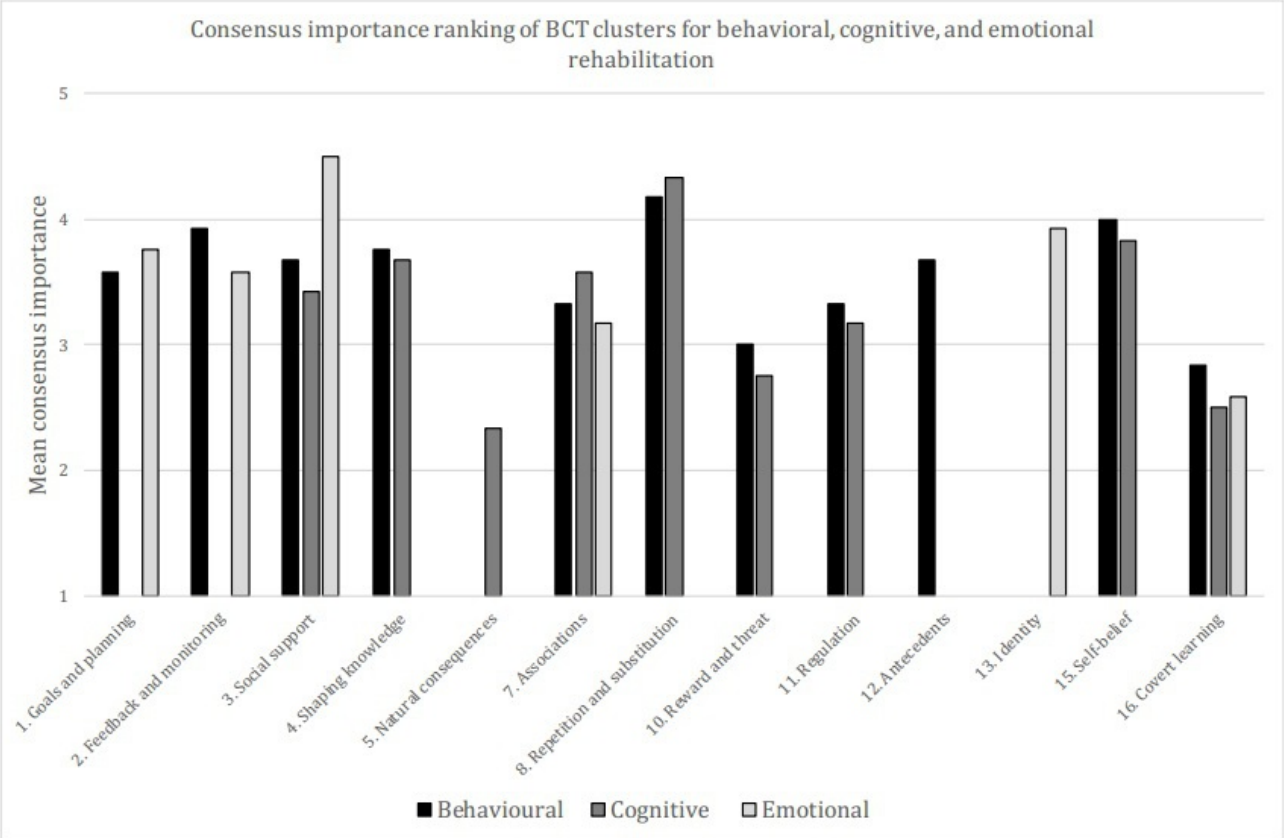
Rank	BCT cluster	Mean (SD)
1	3. Social support	4.50 (0.80)[a]
2	13. Identity	3.92 (0.90)[a]
3	1. Goals and planning	3.75 (0.87)
4	2. Feedback and monitoring	3.58 (0.79)
5	7. Associations	3.17 (0.94)[b]
6	16. Covert learning	2.58 (0.90)[b]

[a]The 3 most relevant.

[b]The 3 least relevant.

XSL•FO
RenderX

Figure 1. Consensus importance ranking of BCT clusters for behavioral, cognitive, and emotional rehabilitation. Clusters are displayed by number. Only those clusters that reached consensus are displayed for each aspect. BCT: behavior change technique.



Discussion

Principal Findings

Particularly in the context of the aging population, stroke rehabilitation is a high priority in health care. Yet, while the BCTs in general have received significant attention [18,33], the differences in importance introduced by long-term, multifaceted (behavioral, cognitive, and emotional) rehabilitation have not been widely studied. This study thus takes an important first step by identifying consensus on the importance of 16 BCT clusters for long-term stroke rehabilitation in general and with respect to the behavioral, cognitive, and emotional aspects of stroke rehabilitation, as detailed in the BCTs in Health Care section. Hence, our results directly address our RQ: Which BCT clusters are the most relevant for long-term stroke rehabilitation in a home setting in relation to treating its behavioral, cognitive, and emotional aspects? Specifically, they highlight 3 main insights for the prioritization of BCTs for long-term behavior change outcomes measured by adherence to treatment.

First, our study further confirms the importance of repetition and substitution and feedback and monitoring as key BCT clusters in this context. This aligns with the ranking of Vestjens et al [33], who highlight repetition and substitution as important for complex interventions in general, as well as that of O'Neill et al [46], who highlight feedback and monitoring as important for preventing smoking relapse in the long term. Further, these clusters were also ranked with social support, aligning with wider discussions of the synergy between these BCT clusters [34], as well as practical implementations via, for example, activity monitors [47]. Notably, both our ranking and the long-term focused part of our study differ substantially from the short-term focused rankings reported by O'Neill et al [46]. As such, our results contribute to a growing body of evidence that long-term behavior change emphasizes specific BCTs, which can vary substantially depending on the specific health situation or aspect of care (behavioral, cognitive, and emotional). Further, our results respond to the barriers identified by Jack et al [8] by directly highlighting BCTs that address low levels of self-efficacy, helplessness, and poor social support. In addition, they align with Campbell et al [15] and Bonnechre et al [16], and work in feedback and monitoring can provide a proxy for a physiotherapist's presence, contact, and supervision of the patient. This provides an important insight for practitioners by pointing to potentially valuable long-term BCT clusters as well as emphasizing the need for caution when evaluating BCT rankings derived with respect to short-term interventions.

Second, our results deprioritize both shaping knowledge and natural consequences, including BCTs such as instructions, information about antecedents and health, social, or emotional consequences, and anticipated regret. This also aligns with the work of Vestjens et al [33] and complements the focus of Nieuwenhuijsen et al [31] on the need for contextual sensitivity in effective interventions. This is a notable finding, as such information-focused BCTs form a major, widely used component in many interventions in practice. For example, Guay et al [48] note that 5 of 12 studies included in a systematic review build on some combination of instructions to guide

XSL•FO
RenderX

behavior change. Thus, while information is a critical component in any intervention, especially in this context, practitioners should consider these BCT clusters as supports to more direct, person-focused BCT clusters, such as feedback and monitoring or repetition and substitution.

Finally, linking to focusing on the most relevant BCTs for treating the behavioral, cognitive, and emotional aspects of rehabilitation, we note 2 important differences evident in the emotional rehabilitation results (Table 6). In this aspect, social support and identity become key BCT clusters, in contrast to Tables 3-5, as well as with the long-term ranking of O'Neill et al [46], where identity-related techniques were low ranked and social support was not included. In addition, there were much lower levels of consensus regarding the emotional aspect. This corresponds to the work of Willems et al [34], who also identified a lack of consensus with respect to "stress management or emotional control training," while emotional aspects were also low ranked by both Vestjens et al [33] and O'Neill et al [46]. This suggests the need for further study of how the emotional aspect of stroke rehabilitation might be supported and how it differs from the more traditional focus on behavioral and cognitive aspects. The differences highlight the need for particular attention by practitioners on this aspect, as specific BCTs are required to support emotional rehabilitation, and that due to the lack of consensus in this area, careful monitoring is needed in order to evaluate their impact, and potential interaction with any behavioral or cognitive focused interventions.

Implications and Recommendations

Our work has several implications for patients and clinicians dealing with stroke recovery. The most salient of these is acknowledging the need for dedicated attention to and support for the emotional aspects of stroke rehabilitation, distinct from behavioral and cognitive aspects. Here, there is a need to maintain a positive, progress-oriented approach, where social support and identity-forming interventions are critical. Further, there is also a more general need to ensure adequate social structures around patients during rehabilitation, going beyond the immediate behavioral and cognitive aspects of recovery. One way to support this is through the design of a social-cooperative platform that allows the patient to track their progress, such as improvements in range of motion, enables therapists to monitor and intervene in the patient's home training, and allows relatives and other stakeholders to support the patient's training efforts on a day-to-day basis. An example of such a system has been proposed in the recent work of Lauer-Schmaltz et al [49] on human digital twins in rehabilitation.

These results also point to several issues to consider during intervention design and clinical implementation. First, there is a need for dedicated long-term strategies to be put in place based on consideration of both long-term behavioral results and relevant BCTs but also effective design approaches for maintaining monitoring and support in the long term. Second, while information is an essential component in supporting understanding, it should not be seen as a major intervention in and of itself. Attention should be paid to coordinating multiple targeted interventions addressing each of the behavioral, cognitive, and emotional aspects. Here, care must be taken that each aspect is both individually supported and that the overall set of support is mutually reinforcing and manageable across aspects. Further, there should be provision for monitoring emotional aspects in addition to more typical behavioral and cognitive monitoring. These aspects may be addressed through, for instance, ongoing dialogue with the patients about their intrinsic motivations for rehabilitation, such as resuming previous work or re-engaging in daily activities like cooking, gardening, or sports.

Limitations

The main limitation of this work is the use of an expert panel rather than patient or intervention data to prioritize the BCTs and to solicit patients' perspectives on the potential impact of the results. Here, the Delphi method was used due to the general lack of access to structured data on the efficacy of BCTs in stroke rehabilitation as well as the general challenge in understanding the specific requirements of long-term behavior change. In this context, the Delphi method provides a robust means of establishing expert consensus. Second, 12 experts were involved in the Delphi survey, and while this is an acceptable number for the Delphi method and in line with prior studies, it places a focus on reaching consensus between diverse experts [38,41,42]. As such, we recruited experts from different fields of specialization, professional background, and nationalities. Thus, while further development of the panel might reveal distinctions between subgroups as in the work of Willems et al [34], the current panel provides a basis for the results reported here as well as directions for further study. Finally, while there is debate as to the required level of consensus for Delphi results, we followed prior guidance in considering BCT clusters to have reached consensus when IQR ≤1 and percentage ≥50% [35,44,50]. Notably, this highlighted a distinct lack of consensus with respect to the emotional aspect of stroke rehabilitation. This is despite this aspect being understood as important by all experts. Thus, while the included results build on prior guidance, there is a need for further study of the exact sources and nature of disagreement regarding the emotional aspect of stroke rehabilitation as well as to explore this aspect more concretely in practice via focused feasibility studies [29].

Conclusions

This international expert panel study using a 2-round Delphi survey ranked, for the first time, the importance of 16 BCT clusters for long-term stroke rehabilitation. The process yielded several new insights highlighting differences in BCT importance between general rehabilitation and rehabilitation specifically focused on the behavioral, cognitive, and emotional aspects of stroke recovery. In particular, we identified a low level of consensus with respect to the emotional aspect but also a distinctly different BCT prioritization in comparison to the other aspects, as such, enabling more effective intervention mapping and guiding the design of multilevel rehabilitation and health promotion interventions and implementation strategies [24]. The study results provide decision support for targeted and tailored design of complex interventions for long-term behavior change and treatment adherence. This provides a first but

XSL•FO
RenderX

important step toward unlocking the prioritization of BCTs for long-term intervention contexts such as stroke rehabilitation.

Acknowledgments

The authors thank the study participants for their time and expertise. Each step of the research has solely been conducted by human authors, and no generative artificial intelligence was used. This study was conducted as a part of the ReHyb project: Rehabilitation Based on Hybrid Neuroprosthesis, funded by the Horizon 2020 Framework Program of the European Union (grant agreement ID 871767).

Authors' Contributions

AEW conducted conceptualization, data curation, formal analysis, investigation, methodology, project administration, validation, and visualization and wrote the original draft. PC conducted conceptualization, data curation, formal analysis, funding acquisition, methodology, project administration, supervision, validation, and visualization, wrote the original draft, and reviewed and edited the manuscript. AM conducted conceptualization, data curation, funding acquisition, methodology, project administration, supervision, and validation and reviewed and edited the manuscript. JPH conducted funding acquisition, methodology, and validation and reviewed and edited the manuscript.

Conflicts of Interest

None declared.

References

1. WHO MONICA Project Principal Investigators. The World Health Organization MONICA project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. J Clin Epidemiol 1988;41(2):105-114. [doi: 10.1016/0895-4356(88)90084-4]
2. Langhorne P, Bernhardt J, Kwakkel G. Stroke rehabilitation. Lancet 2011 May 14;377(9778):1693-1702. [doi: 10.1016/S0140-6736(11)60325-5] [Medline: 21571152]
3. I. Kneebone I, B. Lincoln N. Psychological problems after stroke and their management: state of knowledge. NM 2012;03(01):83-89. [doi: 10.4236/nm.2012.31013]
4. Bates B, Choi JY, Duncan PW, et al. Veterans Affairs/Department of Defense Clinical Practice Guideline for the management of adult stroke rehabilitation care: executive summary. Stroke 2005 Sep;36(9):2049-2056. [doi: 10.1161/01.STR.0000180432.73724.AD] [Medline: 16120847]
5. Wafa HA, Wolfe CDA, Emmett E, Roth GA, Johnson CO, Wang Y. Burden of stroke in Europe: thirty-year projections of incidence, prevalence, deaths, and disability-adjusted life years. Stroke 2020 Aug;51(8):2418-2427. [doi: 10.1161/STROKEAHA.120.029606] [Medline: 32646325]
6. Wolfe CDA. The impact of stroke. Br Med Bull 2000;56(2):275-286. [doi: 10.1258/0007142001903120] [Medline: 11092079]
7. Johnson W, Onuma O, Owolabi M, Sachdev S. Stroke: a global response is needed. Bull World Health Organ 2016 Sep 1;94(9):634-634A. [doi: 10.2471/BLT.16.181636] [Medline: 27708464]
8. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. Man Ther 2010 Jun;15(3):220-228. [doi: 10.1016/j.math.2009.12.004] [Medline: 20163979]
9. Nordin C, Michaelson P, Eriksson MK, Gard G. It's about me: patients' experiences of patient participation in the web behavior change program for activity in combination with multimodal pain rehabilitation. J Med Internet Res 2017 Jan 18;19(1):e22. [doi: 10.2196/jmir.5970] [Medline: 28100440]
10. Teasell R, Mehta S, Pereira S, et al. Time to rethink long-term rehabilitation management of stroke patients. Top Stroke Rehabil 2012;19(6):457-462. [doi: 10.1310/tsr1906-457] [Medline: 23192711]
11. Schmidt R, Stenger K. Behavioral brittleness: the case for strategic behavioral public policy. Behav Public Policy 2024 Apr;8(2):212-237. [doi: 10.1017/bpp.2021.16]
12. Liu S, Hodgson C, Zbib AM, Payne AYM, Nolan RP. The effectiveness of loyalty rewards to promote the use of an internet-based heart health program. J Med Internet Res 2014 Jul 2;16(7):e163. [doi: 10.2196/jmir.3458] [Medline: 24989982]
13. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664]
14. Bassett S. The assessment of patient adherence to physiotherapy rehabilitation. New Zeal J Physiother 2003;31(2):60-66.
15. Campbell R, Evans M, Tucker M, Quilty B, Dieppe P, Donovan JL. Why don't patients do their exercises? Understanding non-compliance with physiotherapy in patients with osteoarthritis of the knee. J Epidemiol Community Health 2001 Feb;55(2):132-138. [doi: 10.1136/jech.55.2.132] [Medline: 11154253]

XSL•FO
RenderX

16. Bonnechre B, Jansen B, Omelina L, Jan S. Do patients perform their exercises at home and why (not)? A survey on patients' habits during rehabilitation exercises. ULUTAS Med J 2016;2(1):41. [doi: 10.5455/umj.20160210060312]
17. Boulton-Lewis GM, Smith DJH, McCrindle AR, Burnett PC, Campbell KJ. Secondary teachers' conceptions of teaching and learning. Learn Instr 2001 Feb;11(1):35-51. [doi: 10.1016/S0959-4752(00)00014-1]
18. 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]
19. Babalola O. Consumers and their demand for healthcare. J Health Med Econ 2017;03(1):3-5. [doi: 10.21767/2471-9927.100032]
20. Chauhan BF, Jeyaraman MM, Mann AS, et al. Behavior change interventions and policies influencing primary healthcare professionals' practice—an overview of reviews. Implement Sci 2017 Jan 5;12(1):3. [doi: 10.1186/s13012-016-0538-8] [Medline: 28057024]
21. Colquhoun HL, Squires JE, Kolehmainen N, Fraser C, Grimshaw JM. Methods for designing interventions to change healthcare professionals' behaviour: a systematic review. Implement Sci 2017 Mar 4;12(1):30. [doi: 10.1186/s13012-017-0560-5] [Medline: 28259168]
22. Beckie TM. A behavior change intervention for women in cardiac rehabilitation. J Cardiovasc Nurs 2006;21(2):146-153. [doi: 10.1097/00005082-200603000-00013] [Medline: 16601534]
23. Hagger MS, Protogerou C. Self‐determination theory and autonomy support to change healthcare behavior. In: Hadler A, Sutton S, Osterberg L, editors. The Wiley Handbook of Healthcare Treatment Engagement: Theory, Research, and Clinical Practice: Wiley Blackwell; 2020:141-158. [doi: 10.1002/9781119129530]
24. Wong AWK, Fong MWM, Munsell EGS, et al. Using intervention mapping and behavior change techniques to develop a digital intervention for self-management in stroke: development study. JMIR Hum Factors 2023 Jul 24;10:e45099. [doi: 10.2196/45099] [Medline: 37486748]
25. Cash PJ, Hartlev CG, Durazo CB. Behavioural design: a process for integrating behaviour change and design. Des Stud 2017 Jan;48:96-128. [doi: 10.1016/j.destud.2016.10.001]
26. Khadilkar PR, Cash P. Understanding behavioural design: barriers and enablers. J Eng Des 2020 Oct 2;31(10):508-529. [doi: 10.1080/09544828.2020.1836611]
27. Fogg BJ. The behavior grid: 35 ways behavior can change. Presented at: Proceedings of the 4th International Conference on Persuasive Technology; Apr 26-29, 2009; Claremont, CA, United States URL: http://dl.acm.org/citation.cfm?id=1542001 [accessed 2025-03-14]
28. O'Brien MK, Shawen N, Mummidisetty CK, et al. Activity recognition for persons with stroke using mobile phone technology: toward improved performance in a home setting. J Med Internet Res 2017 May 25;19(5):e184. [doi: 10.2196/jmir.7385] [Medline: 28546137]
29. Skivington K, Matthews L, Simpson SA, et al. Framework for the development and evaluation of complex interventions: gap analysis, workshop and consultation-informed update. Health Technol Assess 2021;25(57):1-132. [doi: 10.3310/hta25570]
30. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232. [doi: 10.2196/jmir.7126] [Medline: 28663162]
31. Nieuwenhuijsen ER, Zemper E, Miner KR, Epstein M. Health behavior change models and theories: contributions to rehabilitation. Disabil Rehabil 2006 Mar 15;28(5):245-256. [doi: 10.1080/09638280500197743] [Medline: 16492619]
32. Cash P, Khadilkar P, Jensen J, Dusterdich C, Mugge R. Designing behaviour change: a behavioural problem/solution (BPS) matrix. Int J Des 2020;14(2):65-83 [FREE Full text]
33. Vestjens L, Kempen G, Crutzen R, Kok G, Zijlstra GAR. Promising behavior change techniques in a multicomponent intervention to reduce concerns about falls in old age: a Delphi study. Health Educ Res 2015 Apr;30(2):309-322. [doi: 10.1093/her/cyv003] [Medline: 25753146]
34. Willems M, Waninge A, de Jong J, Hilgenkamp TIM, van der Schans CP. Exploration of suitable behaviour change techniques for lifestyle change in individuals with mild intellectual disabilities: a Delphi study. J Appl Res Intellect Disabil 2019 May;32(3):543-557. [doi: 10.1111/jar.12548] [Medline: 30561887]
35. Yip M, Ackery A, Jamieson T, Mehta S. The priorities of end users of emergency department electronic health records: modified Delphi study. JMIR Hum Factors 2023 Mar 10;10:e43103. [doi: 10.2196/43103] [Medline: 36897633]
36. Okoli C, Pawlowski SD. The Delphi method as a research tool: an example, design considerations and applications. Inform Manag 2004 Dec;42(1):15-29. [doi: 10.1016/j.im.2003.11.002]
37. Rowe G, Wright G, Bolger F. Delphi: a reevaluation of research and theory. Technol Forecast Soc Change 1991 May;39(3):235-251. [doi: 10.1016/0040-1625(91)90039-I]
38. Garvin LA, Simon SR. Prioritizing measures of digital patient engagement: a Delphi expert panel study. J Med Internet Res 2017 May 26;19(5):e182. [doi: 10.2196/jmir.4778] [Medline: 28550008]
39. Nasa P, Jain R, Juneja D. Delphi methodology in healthcare research: how to decide its appropriateness. World J Methodol 2021 Jul 20;11(4):116-129. [doi: 10.5662/wjm.v11.i4.116] [Medline: 34322364]

XSL•FO
RenderX

40. Sumsion T. The Delphi technique: an adaptive research tool. Br J Occup Ther 1998 Apr;61(4):153-156. [doi: 10.1177/030802269806100403]
41. Clayton MJ. Delphi: a technique to harness expert opinion for critical decision‐making tasks in education. Educ Psychol (Lond) 1997 Dec;17(4):373-386. [doi: 10.1080/0144341970170401]
42. Humphrey-Murto S, Wood TJ, Gonsalves C, Mascioli K, Varpio L. The Delphi method. Acad Med 2020;95(1):168-168. [doi: 10.1097/ACM.0000000000002887]
43. Overview of mandatory reporting. Danish Research Ethics Committees. 2024. URL: https://researchethics.dk/information-for-researchers/overview-of-mandatory-reporting [accessed 2025-05-25]
44. Gracht HA. Consensus measurement in Delphi studies. review and implications for future quality assurance. Technol Forecast Soc Change 2012;79(8):1525-1536. [doi: 10.1016/j.techfore.2012.04.013]
45. Raskin MS. The Delphi study in field instruction revisited: expert consensus on issues and research priorities. J Soc Work Educ 1994 Jan;30(1):75-89. [doi: 10.1080/10437797.1994.10672215]
46. O'Neill N, Dogar O, Jawad M, Kellar I, Kanaan M, Siddiqi K. Which behavior change techniques may help waterpipe smokers to quit? An expert consensus using a modified Delphi technique. Nicotine Tob Res 2018 Jan 5;20(2):154-160. [doi: 10.1093/ntr/ntw297] [Medline: 27798091]
47. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192. [doi: 10.2196/jmir.3469] [Medline: 25131661]
48. Guay C, Auger C, Demers L, et al. Components and outcomes of internet-based interventions for caregivers of older adults: systematic review. J Med Internet Res 2017 Sep 19;19(9):e313. [doi: 10.2196/jmir.7896] [Medline: 28928109]
49. Lauer-Schmaltz MW, Cash P, Paulin Hansen J, Das N. Human digital twins in rehabilitation: a case study on exoskeleton and serious-game-based stroke rehabilitation using the ETHICA methodology. IEEE Access 2024;12:180968-180991. [doi: 10.1109/ACCESS.2024.3508029]
50. Buss VH, Varnfield M, Harris M, Barr M. A mobile app for prevention of cardiovascular disease and type 2 diabetes mellitus: development and usability study. JMIR Hum Factors 2022 May 10;9(2):e35065. [doi: 10.2196/35065] [Medline: 35536603]

Abbreviations

BCT: behavior change technique
BCTT: behavior change technique taxonomy
RQ: research question

Edited by TDA Cardoso; submitted 04.04.24; peer-reviewed by F Ferraro, S Sheraz; revised version received 25.02.25; accepted 28.02.25; published 07.04.25.

Please cite as:
Wróbel AE, Cash P, Maier A, Paulin Hansen J
Determining the Prioritization of Behavior Change Techniques for Long-Term Stroke Rehabilitation: Delphi Survey Study
Interact J Med Res 2025;14:e59172
URL: https://www.i-jmr.org/2025/1/e59172
doi:10.2196/59172

© Agata Ewa Wróbel, Philip Cash, Anja Maier, John Paulin Hansen. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 7.4.2025. 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

Community Comfort With Automatic Sharing of Race, Ethnicity, and Language Data Between Health Care Settings: Cross-Sectional Study

Noah Brazer[1]; Baylah Tessier-Sherman[1], MPH; Deron Galusha[1], MS; Sakinah C Suttiratana[1], MBA, MPH, PhD; Corrine Liu[1], MPH; Katherine K Kim[2], MBA, MPH, PhD; Mark E Abraham[3], MPH; Marcella Nunez-Smith[1], MHS, MD; Karen H Wang[1], MHS, MD

[1]Department of Internal Medicine, Yale School of Medicine, 100 Church Street South, Suite A200, New Haven, CT, United States

[2]Department of Public Health Sciences, School of Medicine, University of California Davis, Sacramento, CA, United States

[3]DataHaven, New Haven, CT, United States

Corresponding Author:
Karen H Wang, MHS, MD
Department of Internal Medicine, Yale School of Medicine, 100 Church Street South, Suite A200, New Haven, CT, United States

Abstract

Background: Little is known regarding patient attitudes toward automatic sharing of race, ethnicity, and language (REL) data in health care settings despite the universal practice of data sharing across health care institutions and providers.

Objective: This study aims to assess public comfort with disclosing and automatically sharing REL data in health care settings and understand the social factors associated with these attitudes.

Methods: Using the 2022 DataHaven Community Wellbeing Survey from 1196 adult Connecticut residents, we examined factors associated with public comfort with disclosing and automatically sharing REL data across health care settings. We generated unadjusted and adjusted logistic models to examine associations between factors and responses to the data-sharing questions.

Results: Most residents surveyed were White (n=873, 73%), followed by African American or Black (n=167, 14%), Asian or Native Hawaiian or other Pacific Islander (n=31, 2.6%), multiracial (n=31, 2.6%), and American Indian or Alaska Native (n=12, 1%). The majority of respondents were not Hispanic or Latino (n=1051, 87.9%). More than half of respondents reported excellent or very good self-rated health (SRH; n=635, 53.1%), and most participants reported almost always trusting their health care provider (n=939, 78.5%). Most participants reported being willing to share race and ethnicity data at a hospital or clinic (n=1008, 84.3%) and REL data automatically (n=947, 79.2%) in health care settings. Hispanic or Latino (adjusted odds ratio [AOR] 0.049, 95% CI 0.25-0.94) and multiracial (AOR 0.32, 95% CI 0.14-0.76) respondents were less likely to be willing to disclose race and ethnicity data compared to those who were not Hispanic or Latino and who were White, respectively. Individuals who sometimes trust health care providers (AOR 0.57, 95% CI 0.35-0.94) or rarely/never (AOR 0.35, 95% CI 0.15-0.85) were less likely to be willing to disclose race and ethnicity data than those who almost always trust health care providers. African American or Black (AOR 0.46, 95% CI 0.29-0.72) and American Indian or Alaska Native (AOR 0.18, 95% CI 0.04-0.75) individuals were less likely to be willing to share REL data automatically than White individuals. Those who sometimes trust health care providers (AOR 0.48, 95% CI 0.31-0.74) or rarely/never trust health care providers (AOR 0.25, 95% CI 0.11-0.56) were less likely to be willing to share REL data automatically than those who almost always trust health care providers. Those with poor/fair SRH versus very good/excellent SRH were less likely to be willing to share REL data automatically (AOR 0.54, 95% CI 0.34-0.85).

Conclusions: Racial and ethnic identity, SRH, and trust in health care providers affect willingness to share REL information with providers and other health systems.

(Interact J Med Res 2025;14:e67288) doi:10.2196/67288

KEYWORDS

health care disparities; health care justice; data sharing; data autonomy; race; ethnicity; disparity; race, ethnicity, and language; racial identity; ethnic identity; REL

Introduction

Patient data collection pertaining to race, ethnicity, and language (REL) was first mandated in the United States as part of the Civil Rights Act of 1964 as a means to document race-based discrimination in health care settings and drive institutional and social change [1]. This groundbreaking legislative change was one of the first steps to improve health care disparities among

XSL•FO
RenderX

racially and ethnically minoritized people who face marginalization because of systemic oppression. In recent years, REL data collection has become digitized following the passing of the Health Information and Technology for Economic and Clinical Health Act (HITECH) in 2009, which used monetary incentives to encourage the use of electronic health record (EHR) systems [2]. Despite being codified as a federal mandate, REL data collection for patients is not always implemented systematically in health care settings, nor is it always accurate [3]. For example, during the initial height of the COVID-19 pandemic, race and ethnicity data were missing from as many as 56% of reported cases at certain points in time for various public health surveillance data [4]. As a result, health care and public health systems and state government agencies have invested in ensuring improved REL data collection in health [5].

Though race and ethnicity data collection in health care systems is the main mechanism to examine quality of care across diverse populations based on racialized identities, there have been long-standing challenges to consistent data collection despite various improvement efforts [6]. These challenges include concern among clinical staff about asking patients about their information and a lack of training for clinical staff on the importance of data collection. Furthermore, if patients do not have a clear understanding of why these data are being collected and whether the data will be used outside of a clinical setting for research purposes, they may feel less comfortable disclosing information. Efforts to overcome these obstacles to data collection have included reassuring patients about the confidentiality of their data, as well as introducing scripts and training for clinical staff. Beyond individual system-level quality improvement efforts, health systems have been using and sharing this information with other systems across health care settings, which can help standardize data collection practices [7]. While data sharing across health systems and organizations serves as both a mechanism to support public health and streamline care for patients, it may occur without patients' knowledge or explicit consent [1,8-10]. Several prior studies have specifically examined patients' preferences regarding sharing of health data with other clinical providers, finding that patients prefer to be asked permission before having their data shared electronically outside of emergency contexts and that patients who are unwilling to share their data automatically with other health care providers cited concerns about potential breaches in security [11-13]. Although these studies focused on the automatic sharing of general health data, little is known about how patients in the United States feel about sharing specific aspects of their medical records, such as REL data, that are embedded in their EHR [14,15]. One study found no significant difference in patients' willingness to share general health data automatically based on their race and ethnicity; however, this study did not specifically investigate attitudes toward automatic sharing of REL data [16].

This study aims to understand how adults aged 18 and up in Connecticut feel about having their REL data collected and automatically shared with health care systems and whether there are differences in their preferences associated with their racial and ethnic identities, as well as other sociodemographic factors.

Methods

Study Design

This cross-sectional study involved analysis of data from the 2022 DataHaven Community Wellbeing Survey [17]. The DataHaven Community Wellbeing Survey is conducted on an ongoing basis in Connecticut and is designed to measure well-being and quality of life at the statewide, town, and zip code level. The cost of fielding the survey is shared across more than 100 public and private agencies, including all of the state's acute care hospitals, large community foundations, and major cities, as well as many agencies representing rural and suburban communities.

Setting

Survey data were collected by the Siena College Research Institute between August 1 and August 28, 2022. Respondents to the 2022 DataHaven Community Wellbeing Survey were contacted via cell phone and landline or were part of an online exchange. The landline sample included both listed and unlisted telephone numbers, using random digit dialing. The cell phone sample was drawn from dedicated wireless telephone exchanges within Connecticut. Dynata's Wireless LITe database was utilized, which enabled targeting of cell phone samples by region or zip code. The online sample was provided by Lucid, a market research platform that runs an online exchange for survey respondents. The samples drawn from this exchange matched a set of demographic quotas on age, gender, and region. Respondents were sent from Lucid directly to the survey software operated by the Siena College Research Institute. All respondents taking the survey online completed an attention check before and during the survey to ensure proper attention was being paid throughout. The cell phone and landline interviews were conducted in English or Spanish, while online surveys were conducted in English. All respondents were screened for residence in the state of Connecticut.

Participants

The survey sample included 1196 respondents randomly selected from Connecticut residents, aged 18 years or older, to inform state-level estimations of resident well-being, equity, and quality of life, as previously described [17]. The only inclusion criterion for the survey was being a Connecticut resident aged 18 years or older. Among the respondents, 727 (60.8%) completed the survey on a cell phone, 321 (26.8%) completed it on a landline, and 148 (12.4%) completed it online. Respondents represented all 169 Connecticut towns. We followed the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for a cross-sectional study for reporting [18].

Assessments and Data Sources

The outcome variables for this study were delevoped from 2 questions that were added to the 2022 DataHaven Community Wellbeing Survey. The first question was "Patients are often asked their race and ethnicity at a hospital or clinic. Are you OK sharing this information at a hospital or clinic?" Response options were "Yes" or "No." The second question was "If you were offered the choice to have your race/ethnicity/language

XSL•FO
RenderX

information automatically shared electronically with the different places where you receive medical care, how likely would you be to agree to it?" This question was adapted from a prior study [11]. Response options fell along a 4-point Likert scale from "Very likely" to "Very unlikely." "Very likely" and "Somewhat likely" were collapsed into one category, while "Somewhat unlikely" and "Very unlikely" were collapsed into one category.

The covariates for this study were race and ethnicity, gender, age, primary language, reporting of chronic disease, self-rated health (SRH), trust in health care provider, and prior experience with discrimination in health care settings. We selected these covariates as prior research suggests they may influence the relationship between a person and their comfort with sharing personal data for either research or health care use [12,19-21]. Race categories were "African American or Black," "American Indian or Alaska Native," "Asian," "Native Hawaiian or Other Pacific Islander," "White," and "other/something else." Ethnicity categories were "Hispanic or Latino" and "not Hispanic or Latino." Asian and Native Hawaiian or other Pacific Islander were collapsed into one category because of small sample size. Respondents who selected more than one race category were placed in the "Multiracial" category. Chronic disease information was collected by asking respondents about the following conditions: high blood pressure or hypertension; diabetes; heart attack, also called myocardial infarction; angina or coronary heart disease; stroke; and asthma. Respondents who indicated they had any of the conditions were grouped into one category. Respondents with "Excellent" or "Very good" SRH were grouped into one category, while those with "Fair" or "Poor" SRH were grouped into one category. Trust in health care providers was collected by asking how often the following statement was true: "I trust that my health care provider is trying to do what is best for me." Response options were "Almost always," "Sometimes," "Rarely," and "Never." "Rarely" and "Never" were collapsed into one category. Prior experience of discrimination in health care settings was collected by asking, "When seeking health care, have you ever been treated with less respect or received services that were not as good as what other people get?"

Study Size

For our analytic approach, we first conducted univariate analysis to describe survey participants who had any data available for analysis (n=1196). We then removed participants who were missing outcome data (n=69) or covariate data (n=134), leaving a final analytic sample of 993.

Data Analysis

We conducted unadjusted and adjusted parsimonious logistic models to compare demographic characteristics and responses to the data-sharing questions and to determine the independent predictors of our outcomes. We entered covariates with a *P* value <.20 from the unadjusted models into the multivariable regression models using backward elimination, retaining covariates with a *P* value <.05. For the multivariate adjusted model analyzing comfort with REL disclosure, the following variables remained in the parsimonious model: race, ethnicity, gender, and trust in health care provider. For the multivariate adjusted model analyzing comfort with REL sharing, the following variables remained in the parsimonious model: race, age, trust in health care provider, self-rated overall health, and any chronic diseases. Correlations between covariates were examined, and no significant collinearity was observed. All analyses were conducted using SAS Version 9.4 (SAS Institute).

Ethical Considerations

This study did not meet the regulatory criteria for human subjects research because the data from the 2022 DataHaven Community Wellbeing Survey comprise a publicly accessible deidentified database, and as such, institutional review board review and approval were not required [22]. During data collection, informed consent was obtained by DataHaven, and identifiable data was removed before the data were made publicly accessible.

Results

Among the sample of 1196 adult Connecticut residents aged 18 years or older who were randomly selected to participate in the 2022 DataHaven Community Wellbeing Survey, half of the respondents identified as female (Table 1). The majority of residents surveyed were White (n=873, 73%), followed by African American or Black (n=167, 14%), Asian or Native Hawaiian or other Pacific Islander (n=31, 2.6%), multiracial (n=31, 2.6%), and American Indian or Alaska Native (n=12, 1%). Most respondents were not Hispanic or Latino (n=1051, 87.9%). This population slightly overrepresented African American or Black, Native Hawaiian or other Pacific Islander, and American Indian or Alaska Native groups while underrepresenting White, Asian, and Hispanic or Latino groups relative to the Connecticut population reported in the most recent census [23]. In the cohort, 22.1% (n=264) of individuals were between the ages of 18 and 39 years. Only a small percentage of the participants did not speak English as their primary language (n=64, 5.4%). The majority of respondents reported excellent or very good SRH (n=635, 53.1%), and more than half reported a chronic disease condition (n=668, 55.9%). Most participants reported almost always trusting their health care provider (n=939, 78.5%), and the majority did not have a prior experience of discrimination in a health care setting (n=990, 82.8%). Most participants were willing to share race and ethnicity data at a hospital or clinic (n=1008, 84.3%) and were willing to share REL data automatically (n=947, 79.2%; Table 1).

XSL•FO
RenderX

Table . Survey participant characteristics and responses (N=1196).

	Participants, n (%)
Race	
African American/Black	167 (13.9)
American Indian/Alaska Native	12 (1.0)
Asian/Native Hawaiian or other Pacific Islander	31 (2.6)
White	873 (72.9)
Multiracial	31 (2.6)
Missing	82 (6.9)
Ethnicity	
Hispanic or Latino	126 (10.5)
Not Hispanic or Latino	1051 (87.9)
Missing	19 (1.6)
Age (years)	
18 - 39	264 (22.1)
40 - 64	428 (35.8)
65+	455 (38)
Missing	49 (4.1)
Gender	
Female	602 (50.3)
Male	591 (49.4)
Other	3 (0.2)
English as primary language?	
Yes	1122 (93.8)
No	64 (5.4)
Missing	10 (0.8)
Trust health care provider	
Almost always	939 (78.5)
Sometimes	194 (16.2)
Rarely/never	42 (3.5)
Missing	21 (1.8)
Self-rated overall health	
Excellent/very good	635 (53.1)
Good	338 (28.3)
Fair/poor	219 (18.3)
Missing	4 (0.3)
Any chronic disease reported?	
Yes	668 (55.8)
No	519 (43.4)
Missing	9 (0.8)
Prior experience with discrimination	
Yes	170 (14.2)
No	990 (82.8)

XSL•FO
RenderX

	Participants, n (%)
Don’t know	29 (2.4)
Missing	7 (0.6)
Okay sharing race/ethnicity?	
Yes	1008 (84.3)
No	142 (11.9)
Missing	46 (3.8)
Likelihood to automatically share race, ethnicity, and language data	
Very likely/somewhat likely	947 (79.2)
Somewhat unlikely/very unlikely	216 (18.1)
Missing	33 (2.8)

Unadjusted models suggested that the likelihood of being willing to share race and ethnicity data at a hospital or clinic was associated with a respondent’s race, ethnicity, gender, and primary language and how much they trust their health care provider (Table 2). In the multivariate logistic regression model, multiracial participants were significantly less likely to be willing to share their race/ethnicity data compared to White participants (adjusted odds ratio [AOR] 0.32, 95% CI 0.14-0.76; Table 2). Similarly, Hispanic or Latino participants were significantly less likely to be willing to share their race/ethnicity data compared to those who were not Hispanic or Latino (AOR 0.49, 95% CI 0.25-0.94). Female respondents were more likely to be willing to share their race/ethnicity data compared to male respondents (AOR 1.59, 95% CI 1.07-2.41). Individuals who reported they only sometimes (AOR 0.57, 95% CI 0.35-0.94) or rarely/never (AOR 0.35, 95% CI 0.15-0.85) trust that their health care provider has their best interest in mind were less likely to be willing to share race/ethnicity data compared to those who almost always trust their health care provider.

XSL•FO
RenderX

Table . Association between individual characteristics and comfort with disclosing race and ethnicity at hospitals or clinics.

Demographic characteristics	Unadjusted		Multivariate (adjusted)	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Race				
White	Reference	—[a]	Reference	—
African American/Black	0.70 (0.41-1.20)	.19	0.68 (0.39-1.17)	.16
American Indian/Alaska Native	0.37 (0.08-1.79)	.21	0.57 (0.11-3.02)	.51
Asian/Native Hawaiian or other Pacific Islander	0.80 (0.24-2.73)	.72	1.20 (0.33-4.29)	.78
Multiracial	0.29 (0.12-0.67)	.004[b]	0.32 (0.14-0.76)	.01[c]
Ethnicity				
Not Hispanic or Latino	Reference	—	Reference	—
Hispanic or Latino	0.44 (0.24-0.83)	.01[c]	0.49 (0.25-0.94)	.03[c]
Age (years)				
18 - 39	Reference	—	—	—
40 - 64	0.84 (0.49-1.45)	.53	—	—
65+	0.96 (0.56-1.66)	.89	—	—
Gender				
Male	Reference	—	Reference	—
Female	1.55 (1.03-2.34)	.04[c]	1.59 (1.07-2.41)	.03[c]
English as primary language				
Yes	Reference	—	—	—
No	0.41 (0.18-0.93)	.03[c]	—	—
Trust health care provider				
Almost always	Reference	—	—	—
Sometimes	0.53 (0.32-0.86)	.01[c]	0.57 (0.35-0.94)	.03[c]
Rarely/never	0.27 (0.12-0.64)	.003[b]	0.35 (0.15-0.85)	.02[c]
Self-rated overall health				
Excellent/very good	Reference	—	—	—
Good	0.87 (0.55-1.37)	.54	—	—
Fair/poor	0.99 (0.57-1.74)	.98	—	—
Any chronic diseases reported?				
No	Reference	—	—	—
Yes	0.83 (0.55-1.26)	.39	—	—
Prior experience with discrimination in health care?				
No	Reference	—	—	—
Yes	1.24 (0.66-2.33)	.51	—	—

[a]Not applicable.

[b]*P*<.01.

[c]*P*<.05.

In unadjusted models, an individual's likelihood of being willing to share REL data automatically with various health care locations was associated with their race, SRH, and how much they trust their health care provider (Table 3). In the multivariate logistic regression model, African American or Black (AOR 0.46, 95% CI 0.29-0.72) and American Indian or Alaska Native

XSL•FO
RenderX

respondents (AOR 0.18, 95% CI 0.04-0.75) were significantly less likely to be willing to share REL data automatically compared to White respondents (Table 3). Respondents who were aged 40 - 64 years (AOR 0.56, 95% CI 0.34-0.92) and 65+ years (AOR 0.51, 95% CI 0.30-0.87) were significantly less likely to be willing to share REL data than those aged 18 - 39 years. Individuals who reported they only sometimes (AOR 0.48, 95% CI 0.31-0.74) or rarely/never (AOR 0.25, 95% CI 0.11-0.56) trust that their health care provider has their best interest in mind were also significantly less likely to be willing to share REL data automatically than those who almost always trust their health care provider. Respondents with fair/poor SRH were also significantly less likely to be willing to share REL data automatically compared to those with very good/excellent SRH (AOR 0.54, 95% CI 0.34-0.85). In contrast, respondents who reported having any chronic disease were significantly more likely to be willing to share REL data than those who reported not having any chronic disease (AOR 1.58, 95% CI 1.10-2.31).

XSL•FO
RenderX

Table . Association between individual characteristics and willingness to automatically share race, ethnicity, and language data with different health care locations.

Demographic characteristics	Unadjusted		Multivariate (adjusted)	
	Odds ratio (95% CI)	*P* value	Odds ratio (95% CI)	*P* value
Race				
White	Reference	—[a]	Reference	—
African American/Black	0.48 (0.31-0.73)	<.001[b]	0.46 (0.29-0.72)	<.001[b]
American Indian/Alaska Native	0.20 (0.05-0.77)	.02[c]	0.18 (0.04-0.75)	.02[c]
Asian/Native Hawaiian or other Pacific Islander	0.68 (0.25-1.84)	.45	0.57 (0.20-1.65)	.30
Multiracial	0.38 (0.17-0.85)	.02[c]	0.44 (0.19-1.02)	.56
Ethnicity				
Not Hispanic or Latino	Reference	—	—	—
Hispanic or Latino	0.86 (0.46-1.61)	.63	—	—
Age (years)				
18 - 39	Reference	—	Reference	—
40 - 64	0.70 (0.44-1.11)	.13	0.56 (0.34-0.92)	.02[c]
65+	0.87 (0.54-1.38)	.54	0.51 (0.30-0.87)	.01[c]
Gender				
Male	Reference	—	—	—
Female	0.92 (0.66-1.28)	.62	—	—
English as primary language				
Yes	Reference	—	—	—
No	1.28 (0.49-3.33)	.62	—	—
Trust health care provider				
Almost always	Reference	—	—	—
Sometimes	0.49 (0.32-0.74)	<.001[b]	0.48 (0.31-0.74)	<.001[b]
Rarely/never	0.21 (0.10-0.45)	<.001[b]	0.25 (0.11-0.56)	<.001[b]
Self-rated overall health				
Excellent/very good	Reference	—	Reference	—
Good	1.21 (0.79-1.84)	.39	1.18 (0.76-1.84)	.46
Fair/poor	0.54 (0.36-0.81)	.003[d]	0.54 (0.34-0.85)	.008[d]
Any chronic diseases reported?				
No	Reference	—	Reference	—
Yes	1.24 (0.88-1.73)	.22	1.58 (1.10-2.31)	.02[c]
Prior experience with discrimination in health care?				
No	Reference	—	—	—
Yes	0.67(0.43-1.05)	.08	—	—

[a]Not applicable.
[b]*P*<.001.
[c]*P*<.05.
[d]*P*<.01.

XSL•FO
RenderX

Sensitivity analyses showed consistent adjusted results using 2 additional race/ethnicity categorization methods: the common method of grouping race/ethnicity prior to the new 2024 United States Office of Management and Budget race/ethnicity collection guidelines (Tables S1 and S2 in Multimedia Appendix 1) and the rarest classification method to enumerate small populations (Tables S3 and S4 in Multimedia Appendix 1) [24].

Discussion

Principal Results

In general, the majority of Connecticut respondents are comfortable disclosing their race and ethnicity and having their REL data shared automatically. However, racially and ethnically minoritized individuals are less comfortable disclosing race and ethnicity information at hospitals and clinics and having their REL data automatically shared with health institutions where they receive care. As compared to respondents who were White and were not Hispanic or Latino, multiracial and Hispanic or Latino respondents were the least likely to be willing to disclose their race and ethnicity data, respectively, while Black or African American and American Indian or Alaska Native respondents were less likely to be willing to have their REL data automatically shared between medical providers.

Comparison With Prior Work

While previous studies showed no significant difference in respondents' willingness to automatically share general health data between race and ethnicity identity groups [11,16], the results of this study add a more nuanced view of specific data elements within the EHR that certain minoritized individuals feel less inclined to share automatically. It is possible that prior studies saw no significant difference in willingness to automatically share general health data because respondents were not explicitly informed about all the sensitive data that comprised their EHR, such as REL data. Additionally, the previous studies were conducted in California, which has a substantially different representation of race and ethnicity both in the survey samples and the state population than in the current study in Connecticut, with higher representation of Hispanic or Latino and Asian or Native Hawaiian or other Pacific Islander respondents and lower representation of Black or African American respondents. In the United States, Black, Indigenous, and Latino communities have faced discrimination when accessing health care systems, leading to poorer health outcomes and health care delivery than their White counterparts [25,26]. These discriminatory experiences may contribute broadly to diminished comfort with REL data collection and automatic sharing [27].

The finding that lower SRH was independently associated with a decreased likelihood of being willing to share REL data automatically warrants further examination. Past studies have demonstrated a relationship between race and ethnicity and SRH [28-30]. Furthermore, studies investigating the relationship between SRH and likelihood to share data with physicians have shown that those with lower SRH are more likely to share information with their physician, which is consistent with our finding that individuals with chronic disease were more likely to be willing to share REL data [31,32]. However, among our study population, no correlation was found between SRH and chronic disease, suggesting that Connecticut residents may use other metrics to determine their SRH beyond chronic disease. Given the findings of this study, it may be possible that individuals with lower SRH are more comfortable with sharing general health data specifically for treatment purposes but draw the line at sharing REL data, whereas those with chronic disease are comfortable sharing all elements of their data. Follow-up studies could further explore relationships between one's social identities, health status, and sharing of information.

Our finding that trust in health care providers is highly associated with comfort with sharing is consistent with past studies that have established a link between patient trust in their provider and willingness to disclose or share health information [19,33]. Additionally, a previous study found differences in health information seeking behavior of individuals by race and ethnicity, with Black and Hispanic individuals, compared to White individuals, more trusting of health information from media, charitable organizations, and religious organizations and Hispanic individuals, compared to White individuals, less trusting of and willing to seek information from doctors [34]. The results of our study further highlight the importance of trust between patient and provider and reveal that individuals with decreased levels of trust in their health care providers may be hesitant to disclose and share any data, whether health-specific or demographic. Our initial work in this area suggests that trust in health care providers may mediate these relationships between the correlates of interest and outcomes (data not shown). Future studies can explore these relationships.

Limitations

This study had several limitations. Namely, the number of individuals identifying as American Indian or Alaska Native or Native Hawaiian or other Pacific Islander was very low, which affected the statistical analyses. Given that individuals who identify as American Indian or Alaska Native or Native Hawaiian or other Pacific Islander tend to be minoritized in most areas of the United States, a larger cohort would be needed to better understand the nuances of comfort with REL data disclosure and sharing among these communities. Additionally, the 2022 DataHaven Community Wellbeing Survey was only conducted in English or Spanish, which limits representation of communities in Connecticut that do not speak either of these two languages. Further, the population of Connecticut is not representative of that of the United States, and therefore, generalizations made while using these results should be done with caution.

Conclusion

The results of this study have implications for how we operationalize health information exchange across clinical and public health settings in the current discussion of how to most effectively enhance the nationwide health data ecosystem and standardize social and structural drivers of health data [35]. This study suggests that nuances exist in how people think about race and ethnicity data and the sharing of specific information within their health record. People's individual backgrounds likely influence their comfort with sharing sensitive data and their comfort with that data being shared automatically with

XSL•FO
RenderX

other health settings. More work is needed to gauge whether patients have an understanding of what components comprise their health record, which of these components may be shared, and with whom the data are being shared.

Acknowledgments

We thank the team at DataHaven for collecting the data and making it available to us. Generative artificial intelligence was not used in any portion of this manuscript. We acknowledge Dr Daniel Sarpong for assistance with statistical analysis.

This publication was made possible in part by the Genentech Health Equity Innovation Fund and National Institutes of Health/National Library of Medicine (G08LM013801). Funders did not contribute to this work.

Data Availability

The datasets used or analyzed in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

NB, BTS, SCS, and KHW were involved in conceptualization. NB, BTS, DG, SCS, CL, MEA, and KHW were involved in data curation. BTS and DG were involved in formal analysis. MNS and KHW were involved in funding acquisition. KKK, MEA, MNS, and KHW were involved in the investigation. NB, BTS, DG, SCS, KKK, MEA, and KHW were involved in methodology. NB, KKK, MEA, MNS, and KHW were involved in project administration. KKK, MEA, MNS, and KHW were involved in resources. MNS and KHW were involved in supervision. BT and DG were involved in validation. NB, BTS, CL, and KHW were involved in visualization. NB, BTS, CL, and KHW were involved in writing the original draft. All authors were involved in reviewing and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sensitivity analyses comparing two additional race/ethnicity categorization methods.
[DOCX File, 32 KB - i-jmr_v14i1e67288_app1.docx]

References

1. Title VI of the Civil Rights Act of 1964. US Department of Justice. URL: https://www.justice.gov/crt/fcs/TitleVI [accessed 2023-11-23]
2. Health Information Technology for Economic and Clinical Health Act. US Department of Health and Human Services. 2009. URL: https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/enforcementrule/enfifr.pdf [accessed 2023-11-23]
3. Filice CE, Joynt KE. Examining race and ethnicity information in Medicare administrative data. Med Care 2017 Dec;55(12):e170-e176. [doi: 10.1097/MLR.0000000000000608] [Medline: 29135782]
4. Tsai JW, Khazanchi R, Laflamme E. Death by missing data: uncollected racial and ethnic pandemic data will drive inequities for decades to come. STAT. 2023 Jan 30. URL: https://www.statnews.com/2023/01/30/covid-19-missing-data-race-ethnicity-drive-inequities-decades-to-come/ [accessed 2024-04-01]
5. Weissman JS, Adler RR, Betancourt J, et al. How hospitals are addressing the effects of racism: a mixed-methods study of hospital equity officers. Health Aff (Millwood) 2023 Oct;42(10):1402-1410. [doi: 10.1377/hlthaff.2023.00474] [Medline: 37782867]
6. Hasnain-Wynia R, Baker DW. Obtaining data on patient race, ethnicity, and primary language in health care organizations: current challenges and proposed solutions. Health Serv Res 2006 Aug;41(4 Pt 1):1501-1518. [doi: 10.1111/j.1475-6773.2006.00552.x] [Medline: 16899021]
7. Holmgren AJ, Esdar M, Hüsers J, Coutinho-Almeida J. Health information exchange: understanding the policy landscape and future of data interoperability. Yearb Med Inform 2023 Aug;32(1):184-194. [doi: 10.1055/s-0043-1768719] [Medline: 37414031]
8. Kalkman S, van Delden J, Banerjee A, Tyl B, Mostert M, van Thiel G. Patients' and public views and attitudes towards the sharing of health data for research: a narrative review of the empirical evidence. J Med Ethics 2022 Jan;48(1):3-13. [doi: 10.1136/medethics-2019-105651] [Medline: 31719155]
9. Kim J, Kim H, Bell E, et al. Patient perspectives about decisions to share medical data and biospecimens for research. JAMA Netw Open 2019 Aug 2;2(8):e199550. [doi: 10.1001/jamanetworkopen.2019.9550] [Medline: 31433479]
10. Shapiro JS, Mostashari F, Hripcsak G, Soulakis N, Kuperman G. Using health information exchange to improve public health. Am J Public Health 2011 Apr;101(4):616-623. [doi: 10.2105/AJPH.2008.158980] [Medline: 21330598]

XSL•FO
RenderX

11. Kim KK, Joseph JG, Ohno-Machado L. Comparison of consumers' views on electronic data sharing for healthcare and research. J Am Med Inform Assoc 2015 Jul;22(4):821-830. [doi: 10.1093/jamia/ocv014] [Medline: 25829461]
12. Ancker JS, Edwards AM, Miller MC, Kaushal R. Consumer perceptions of electronic health information exchange. Am J Prev Med 2012 Jul;43(1):76-80. [doi: 10.1016/j.amepre.2012.02.027] [Medline: 22704751]
13. Medford-Davis LN, Chang L, Rhodes KV. Health information exchange: what do patients want? Health Informatics J 2017 Dec;23(4):268-278. [doi: 10.1177/1460458216647190] [Medline: 27245671]
14. Stone MA, Redsell SA, Ling JT, Hay AD. Sharing patient data: competing demands of privacy, trust and research in primary care. Br J Gen Pract 2005 Oct;55(519):783-789. [Medline: 16212854]
15. Patil S, Lu H, Saunders CL, Potoglou D, Robinson N. Public preferences for electronic health data storage, access, and sharing—evidence from a pan-European survey. J Am Med Inform Assoc 2016 Nov;23(6):1096-1106. [doi: 10.1093/jamia/ocw012] [Medline: 27107445]
16. Kim KK, Sankar P, Wilson MD, Haynes SC. Factors affecting willingness to share electronic health data among California consumers. BMC Med Ethics 2017 Apr 4;18(1):25. [doi: 10.1186/s12910-017-0185-x] [Medline: 28376801]
17. DataHaven Community Wellbeing Survey. DataHaven. 2023. URL: https://www.ctdatahaven.org/reports/datahaven-community-wellbeing-survey [accessed 2023-11-15]
18. Cuschieri S. The STROBE guidelines. Saudi J Anaesth 2019 Apr;13(Suppl 1):S31-S34. [doi: 10.4103/sja.SJA_543_18] [Medline: 30930717]
19. Iott BE, Campos-Castillo C, Anthony DL. Trust and privacy: how patient trust in providers is related to privacy behaviors and attitudes. AMIA Annu Symp Proc 2019;2019:487-493. [Medline: 32308842]
20. Ancker JS, Witteman HO, Hafeez B, Provencher T, Van de Graaf M, Wei E. The invisible work of personal health information management among people with multiple chronic conditions: qualitative interview study among patients and providers. J Med Internet Res 2015 Jun 4;17(6):e137. [doi: 10.2196/jmir.4381] [Medline: 26043709]
21. Thompson HM. Patient perspectives on gender identity data collection in electronic health records: an analysis of disclosure, privacy, and access to care. Transgend Health 2016;1(1):205-215. [doi: 10.1089/trgh.2016.0007] [Medline: 28861535]
22. Yale University Human Research Protection Program (HRPP): policy and standard operating procedure manual. Yale University: Research Support. 2023 Jul 24. URL: https://research-support.yale.edu/sites/default/files/2025-07/HRPP_Policy_and_Standard_Operating_Procedure_Manual_20230724.pdf [accessed 2025-07-21]
23. QuickFacts: Connecticut. United States Census Bureau. URL: https://www.census.gov/quickfacts/fact/table/CT/ [accessed 2024-05-22]
24. Mays VM, Ponce NA, Washington DL, Cochran SD. Classification of race and ethnicity: implications for public health. Annu Rev Public Health 2003;24:83-110. [doi: 10.1146/annurev.publhealth.24.100901.140927] [Medline: 12668755]
25. Williams DR, Mohammed SA. Discrimination and racial disparities in health: evidence and needed research. J Behav Med 2009 Feb;32(1):20-47. [doi: 10.1007/s10865-008-9185-0] [Medline: 19030981]
26. National healthcare quality and disparities report. Agency for Healthcare Research and Quality. 2022. URL: https://www.ahrq.gov/research/findings/nhqrdr/nhqdr22/index.html [accessed 2023-11-23]
27. Haley JM, Dubay L, Garrett B. Implementing revised federal race/ethnicity data standards won't sufficiently address health inequities. Health Affairs Forefront 2024. [doi: 10.1377/forefront.20240828.438120]
28. Williams DR, Mohammed SA. Racism and health I: pathways and scientific evidence. Am Behav Sci 2013 Aug 1;57(8):1152-1173. [doi: 10.1177/0002764213487340] [Medline: 24347666]
29. Garcia JA, Sanchez GR, Sanchez-Youngman S, Vargas ED, Ybarra VD. Race as lived experience: the impact of multi-dimensional measures of race/ethnicity on the self-reported health status of Latinos. Du Bois Rev 2015;12(2):349-373. [doi: 10.1017/S1742058X15000120] [Medline: 26681972]
30. Macintosh T, Desai MM, Lewis TT, Jones BA, Nunez-Smith M. Socially-assigned race, healthcare discrimination and preventive healthcare services. PLoS ONE 2013;8(5):e64522. [doi: 10.1371/journal.pone.0064522] [Medline: 23704992]
31. Ivanov A, Sharman R, Rao HR. Exploring factors impacting sharing health-tracking records. Health Policy Technol 2015 Sep;4(3):263-276. [doi: 10.1016/j.hlpt.2015.04.008]
32. Kim J, Im E, Kim H. From intention to action: the factors affecting health data sharing intention and action. Int J Med Inform 2023 Jul;175:105071. [doi: 10.1016/j.ijmedinf.2023.105071] [Medline: 37099875]
33. Maiorana A, Steward WT, Koester KA, et al. Trust, confidentiality, and the acceptability of sharing HIV-related patient data: lessons learned from a mixed methods study about health information exchanges. Implement Sci 2012 Apr 19;7:34. [doi: 10.1186/1748-5908-7-34] [Medline: 22515736]
34. Fareed N, Swoboda CM, Jonnalagadda P, Walker DM, Huerta TR. Differences between races in health information seeking and trust over time: evidence from a cross-sectional, pooled analyses of HINTS data. Am J Health Promot 2021 Jan;35(1):84-92. [doi: 10.1177/0890117120934609] [Medline: 32588638]
35. Seidman G, AlKasir A, Ricker K, Lane JT, Zink AB, Williams MA. Regulations and funding to create enterprise architecture for a nationwide health data ecosystem. Am J Public Health 2024 Feb;114(2):209-217. [doi: 10.2105/AJPH.2023.307477] [Medline: 38207252]

XSL•FO
RenderX

Abbreviations

AOR: adjusted odds ratio
EHR: electronic health record
HITECH: Health Information and Technology for Economic and Clinical Health Act
REL: race, ethnicity, and language
SRH: self-rated health
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by TDA Cardoso; submitted 07.10.24; peer-reviewed by F Mouhanna, Y Cuffee; revised version received 28.07.25; accepted 29.07.25; published 06.10.25.

Please cite as:
Brazer N, Tessier-Sherman B, Galusha D, Suttiratana SC, Liu C, Kim KK, Abraham ME, Nunez-Smith M, Wang KH
Community Comfort With Automatic Sharing of Race, Ethnicity, and Language Data Between Health Care Settings: Cross-Sectional Study
Interact J Med Res 2025;14:e67288
URL: https://www.i-jmr.org/2025/1/e67288
doi:10.2196/67288

© Noah Brazer, Baylah Tessier-Sherman, Deron Galusha, Sakinah C Suttiratana, Corrine Liu, Katherine K Kim, Mark E Abraham, Marcella Nunez-Smith, Karen H Wang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 6.10.2025. 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

Evaluating the Safety and Efficacy of a Non–Weight-Centric Approach to Obesity Prevention in Rural and Urban Female Adolescents: Quasi-Experimental Study

Areeg Zuair[1], PhD; Rola Jalloun[2], PhD; Naif Alzahrani[3], PhD; Fahad Alhowaymel[4], PhD; Esraa Merza[3], RN; Bandar Alhumaidi[3], PhD; Mohammad Alahmadi[3], Prof Dr

[1]Community Health Nursing Department, Taibah University, FGPR+CVM Janadah Bin Umayyah Road, Medina, Saudi Arabia

[2]Nutrition and Food Science Department, Taibah University, Medina, Saudi Arabia

[3]Department of Sport Science and Physical Activity, Taibah University, Medina, Saudi Arabia

[4]Department of Nursing Science, Shaqra University, Riyadh, Saudi Arabia

Corresponding Author:
Areeg Zuair, PhD
Community Health Nursing Department, Taibah University, FGPR+CVM Janadah Bin Umayyah Road, Medina, Saudi Arabia

Abstract

Background: Obesity is rising among Saudi adolescents, with rural females particularly underserved due to limited health education and sociocultural barriers. Global guidelines promote non–weight-centric approaches to reduce stigma. The Green Apple program delivers school-based, weight-neutral education, with an added metabolic noncommunicable disease (MNCD) prevention unit. Although previously tested in urban settings, it has not yet been evaluated among rural female students.

Objective: This study aimed to examine the effectiveness of the Green Apple intervention in improving MNCD knowledge and its safety, defined as no adverse psychological (body image discrepancy and disordered eating) or behavioral (sedentary behavior) outcomes.

Methods: This quasi-experimental study included 105 participants from urban and rural schools in Saudi Arabia, with 4 classes assigned to either the intervention group (Green Apple) or the enhanced intervention group (Green Apple+MNCD). Both programs were implemented in female-only classrooms by trained facilitators during regular school hours. The intervention was delivered once per week over 2 consecutive weeks (2 sessions), while the enhanced intervention included an additional third session. Linear mixed-effects models assessed intervention effects across three time points: baseline, postintervention, and 1-month follow-up.

Results: The mean age across participants was 16.42 (SD 0.66) years, with a significant difference between groups: 15.97 (SD 0.41) years in the enhanced intervention group and 17.00 (SD 0.42) years in the intervention group (P<.001). Both interventions significantly improved knowledge across schools. The enhanced intervention (Green Apple+MNCD) group demonstrated an increase of 1.65 (95% CI 0.61-2.70; P<.001) from baseline to follow-up, while the intervention (Green Apple) group showed an increase of 1.26 (95% CI 0.10-2.43; P=.02). However, no significant between-group differences were observed at baseline (mean difference 0.20, SE 0.46; P=.65), postintervention (mean difference 0.79, SE 0.45; P=.08), or follow-up (mean difference 0.73, SE 0.49; P=.13). Although sedentary behavior did not significantly decrease across all schools, a significant reduction was observed in rural schools receiving the Green Apple intervention (–3.12, 95% CI –5.67 to −0.56; P=.02). Regarding safety outcomes, no adverse effects on body image or disordered eating were observed. A significant reduction in disordered eating symptoms was found only in urban schools receiving the enhanced intervention (–0.94, 95% CI –1.61 to –0.26; P=.007).

Conclusions: The Green Apple program is a culturally tailored, weight-neutral intervention that improves metabolic health literacy and reduces sedentary behavior among Saudi female adolescents without harming body image or eating behaviors. By focusing on an underserved subgroup, it addresses a key gap in health promotion and aligns with global calls for stigma-free approaches. Broader regional studies are needed to assess its long-term impact.

(Interact J Med Res 2025;14:e71341) doi:10.2196/71341

KEYWORDS

adolescent obesity; macronutrient education; weight stigma; health literacy; Saudi Arabia; rural

XSL•FO
RenderX

Introduction

As of 2020, over 100 organizations have endorsed the Joint International Consensus Statement or pledged their commitment to eliminating weight stigma [1]. These endorsements come from a diverse range of entities, including scientific societies (eg, American Society for Nutrition, UK Association for the Study of Obesity, and Obesity Australia), academic institutions, professional organizations, medical journals (eg, *Obesity*, *Obesity Reviews*, *Obesity Science and Practice*), and parliamentary groups across the globe [1]. Weight stigma has been linked to anxiety, suicidality, disordered eating, sedentary behavior, and even weight gain [2-4]. Among adolescents, weight stigma is strongly linked to disordered eating, body dissatisfaction, and psychological distress [4]. A comprehensive systematic review by Levinson et al [5] identified consistent associations between experienced, anticipated, and internalized weight stigma and harmful eating behaviors. However, most studies were cross-sectional and based on predominantly White, female samples, limiting generalizability [5]. In addition, school-based obesity prevention programs that emphasize weight loss, BMI reporting, or calorie restriction have been associated with increased body dissatisfaction and disordered eating behaviors among adolescents [5,6]. Physiologically, stigma can activate stress responses, while psychologically, it damages self-esteem and fosters body shame, leading to disordered behaviors [7,8]. These effects underscore the need for weight-neutral interventions that avoid reinforcing weight-based shame [7,8].

In response to the growing awareness of weight stigma, non–weight-centric approaches, which focus on overall health and well-being rather than weight loss, are increasingly being adopted in adolescent school-based obesity interventions [9]. These programs emphasize holistic well-being, body acceptance, intuitive eating, and joyful movement, while tracking outcomes such as blood pressure and psychological health instead of BMI [9]. Evidence suggests that non–weight-centric interventions are associated with improved metabolic health, reduced disordered eating, and greater sustainability of healthy behaviors, while also significantly decreasing internalized weight bias and body dissatisfaction [6,9].

Dugmore et al [6] conducted a meta-analysis comparing weight-neutral and traditional weight loss interventions. Weight-neutral approaches, often guided by Health at Every Size (HAES) principles, showed comparable or superior outcomes for bulimia nervosa, self-esteem, and eating behaviors, without adverse effects on physical health. They also had lower dropout rates and better adherence [6]. However, the studies were few, small, and methodologically inconsistent. Overall, the evidence supports adopting culturally sensitive, weight-neutral interventions that focus on health behaviors rather than weight loss, particularly for vulnerable adolescent populations [6].

However, balancing obesity prevention with avoiding weight-based stigma remains a global challenge, prompting calls for research beyond Western cultural contexts [1,10]. Since perceptions of obesity and weight stigma differ across languages and cultures, it is essential to explore non–weight-centric prevention strategies in non–English-speaking regions [1,10]. Increasing awareness of these cultural and linguistic differences is vital to building inclusive public health narratives that reduce stigma globally [1,10].

In Saudi Arabia, an Arabic-speaking country, adolescent obesity rates are rapidly increasing [11,12]. Recent studies report an obesity and overweight prevalence of 31.6% (2504/7931) among Saudi adolescents [13]. In response, the Saudi government has launched initiatives like the “Strategy to Combat Obesity and Promote Physical Activity in the Arab Countries” and the “RASHAKA Program,” a school-based initiative focused on increasing physical activity and reducing body weight [14,15]. Although no studies have directly measured the psychosocial consequences of these programs in Saudi Arabia, global guidance from the World Obesity Federation cautions that language used in such campaigns — such as “eliminate obesity” or “combat obesity” — can reinforce stigma by portraying individuals with higher weight as problems to be solved [1,10]. To counter this, school-based obesity prevention programs should shift focus from weight reduction to overall well-being, moving away from the “eat less, move more” narrative [10].

A recent review of obesity prevention studies in Saudi Arabia identified only 8 interventions, despite the nation’s high prevalence of adolescent obesity [16-18]. Most studies prioritized weight reduction, emphasizing obesity prevalence and associated health risks and promoting reduced portion sizes, caloric intake reduction, and increased physical activity [16,17]. Even when interventions included components aimed at fostering positive body image and reducing stigma, the primary focus remained on weight reduction and highlighting obesity prevalence and related health risks [19]. Such framing contradicts current guidance, which discourages language that may perpetuate weight stigma or promote disordered weight-control behaviors [10,20]. A 2024 cross-sectional study conducted in Saudi Arabia found that 63% of adolescents (1023/1624) were dissatisfied with their body weight [21]. The results demonstrated significant positive correlations between weight stigma, mental health issues, and BMI, with a statistically significant regional variation.

Despite increasing attention to adolescent health in Saudi Arabia, current initiatives mostly adopt a one-size-fits-all approach that overlooks the distinct needs of rural adolescents. These populations face structural and sociocultural barriers, including limited access to health education, physical activity resources, trained personnel, and specialized health care services, placing them at a disadvantage compared with urban peers [22-24]. Rapid urbanization may also be reshaping rural communities in ways that existing public health strategies have yet to address, underscoring the need for contextually adapted, weight-neutral interventions that promote health without reinforcing stigma or excluding vulnerable subgroups.

Focusing on rural female adolescents is especially important, as this subgroup encounters intersecting barriers related to geography, gender norms, and socioeconomic status [22-24]. Compared with urban girls, they experience limited mobility due to cultural expectations, as well as fewer educational and

XSL•FO
RenderX

recreational opportunities, particularly in relation to structured physical activity [22-24]. These challenges are further intensified by higher poverty rates and lower levels of parental education, which can limit access to health information and preventive care [22,23]. By targeting this underserved population, this study addresses a critical gap in the literature and responds to the urgent need for culturally tailored, equitable health interventions for those most at risk of poor health outcomes. This study is the first to implement and evaluate a culturally adapted, weight-neutral school-based intervention specifically designed for and delivered to Saudi rural female adolescents.

The Green Apple program was intentionally designed as a universal school health initiative aimed at addressing the specific challenges faced by rural and urban female adolescents in Saudi Arabia. By adopting a non–weight-centric, health-focused approach, the program avoids stigmatizing language and instead promotes positive, practical health behaviors relevant to daily life. It delivers culturally sensitive, age-appropriate content in Arabic, ensuring accessibility for students in underresourced rural areas. Importantly, the program emphasizes macronutrient education, metabolic disease prevention, and strength-building exercises that do not require special facilities, making it feasible for rural schools with limited infrastructure. Accordingly, this study was conducted to evaluate the effectiveness of the Green Apple program in enhancing knowledge about MNCDs and sedentary behavior, while also monitoring safety outcomes related to body image and disordered eating symptoms.

Methods

Study Design and Participants

This study employed a quasi-experimental design with 4 parallel arms, incorporating blinding and nonequivalent groups. Of the total, 2 arms received the intervention (1 urban school and 1 rural school), while the other 2 arms received the enhanced intervention (1 urban school and 1 rural school). The study was conducted in 2023 at 1 urban public female high school and 1 rural public female high school in the Medina region of Saudi Arabia. Refer to Checklist 1 for the checklist used to report this study, as recommended by the TREND Group for reporting nonrandomized evaluations of behavioral and public health interventions [25].

The schools were purposefully selected by the Ministry of Education's School Health Department to represent urban and rural populations and ensure administrative feasibility and support. The urban school was in a city center with greater access to health care services and extracurricular health programs, while the rural school served a geographically isolated population with fewer health education resources. In addition, both schools were chosen for their history of successful collaboration with the Ministry of Education, ensuring administrative support and feasibility for program implementation.

Classes within each school were assigned to one of two program versions: (1) enhanced intervention (Green Apple+MNCD): a 3-unit, non–weight-centric program that included content on the prevention of MNCDs and (2) intervention (Green Apple): a 2-unit, non–weight-centric program without MNCD content. A randomized controlled trial (RCT) design was considered during the planning phase; however, logistical, ethical, and administrative constraints necessitated the use of a quasi-experimental design instead. RCT was not feasible due to real-world constraints, including the limited number of schools and classes, especially in rural areas. For example, in the selected rural region, only 1 secondary school for female students was available, and each grade level contained only 1 class. Randomization at the individual or class level would have introduced risk of contamination and was not permissible under school policy. Therefore, a quasi-experimental design was used.

To reduce potential selection bias, first-grade students in both urban and rural schools were assigned to the enhanced intervention condition, while second-grade students were assigned to the intervention condition. This decision was guided by the understanding that second-grade students had completed a Ministry-mandated "Healthy Lifestyles" course and would likely demonstrate higher baseline knowledge. To minimize bias from this difference, the enhanced intervention was deliberately limited to first-grade students to ensure outcome comparability. In addition, a baseline assessment of knowledge revealed no significant differences between the intervention versions. To further strengthen internal validity, linear mixed-effects models were used to adjust for potential confounding factors. These models included time, school type, and intervention type as fixed effects. Although propensity score matching is a common method for reducing selection bias in quasi-experimental studies, it was not implemented due to the limited sample size and the clustered structure of the school-based data [26,27]. Instead, baseline equivalence between groups was assessed, and no significant differences were found in preintervention knowledge scores. While quasi-experimental designs limit the ability to make causal inferences compared with randomized trials, several strategies were used to enhance internal validity. These included grade-level assignment, baseline equivalence testing, and statistical covariate control. Blinding was maintained at the classroom level between intervention groups, as well as during data collection, to reduce performance and detection bias. Nonetheless, potential limitations such as residual confounding or contextual differences across classrooms cannot be fully ruled out. All participants completed self-reported measures at 3 time points: before the intervention (baseline), immediately after the program was delivered (postprogram), and 1 month after the program was delivered (follow-up). Only students who were enrolled in the participating schools, attended the classroom sessions where the intervention was delivered, and provided both informed consent (from guardians) and assent (from students) were included in the analysis. Students who did not provide consent or assent were excluded from data collection and subsequent analysis.

Green Apple Program

Green Apple is an innovative, educational, and universal program aimed at shifting the focus from weight and obesity to the role of liver fatty acids and vascular fat in MNCDs and how lifestyle management can address conditions like diabetes and high blood cholesterol. Unlike traditional programs, Green

XSL•FO
RenderX

Apple deliberately avoids terms like “obesity,” “weight reduction,” “eat less,” and “move more.” Instead, it encourages critical thinking about macronutrients and their relationship to body energy, cardiometabolic disease prevention, and lifestyle management of MNCDs.

The Green Apple program is grounded in 3 complementary theories of behavior change: the health belief model (HBM), social cognitive theory (SCT), and the transtheoretical model (TTM) [28-30]. These models collectively inform the program’s structure and messaging. According to HBM, adolescents are more likely to change their behaviors when they perceive themselves at risk for serious health conditions — such as diabetes, fatty liver, or stroke — and believe that actionable steps can mitigate those risks. SCT emphasizes the importance of self-efficacy and skill-building, which Green Apple fosters through experiential learning, practical lifestyle tips, and positive reinforcement. The TTM supports the idea that behavior change is a staged process, and the Green Apple program is designed to guide students from awareness to long-term action by progressively reinforcing key health concepts across sessions.

Rather than focusing on obesity or weight reduction, the program reframes the discussion around visceral fat and liver fat, highlighting their roles in MNCDs. This framing provides a more neutral, stigma-free way to promote healthy behaviors such as balanced macronutrient intake, strength-building activities, and adherence to MyPlate dietary guidelines [31,32]. The aim is to enhance students’ understanding of metabolic health while improving self-efficacy and motivation for lifelong health habits. In addition, it empowers students with knowledge and practical lifestyle tips to support family members or friends with cardiometabolic diseases without contributing to stigma for those affected by these conditions. To ensure cultural and developmental appropriateness, all materials were delivered in Arabic and tailored to the gender-specific school environment. Sessions were delivered in female-only classrooms by 2 female researchers — one with a PhD in Nursing and the other in Clinical Nutrition — both of whom have experience in teaching school-aged students. The program was delivered once per week over a 2-week period, with each session taking 1 class period (approximately 45 min). Teaching methods included active discussion, visual aids such as culturally adapted Microsoft PowerPoint slides, and non–weight-stigmatizing illustrations.

The Green Apple program was recently evaluated in a pilot study involving 69 urban students (46 boys and 23 girls; mean age 16.18, SD 0.53 y) [33]. The intervention significantly improved MNCD knowledge ($F_{1,64}$=23.45; P<.001), with a large effect size (partial η^2=0.27). Notably, no significant change in students’ body image discrepancy was observed from pre- to postintervention (P=.70), reinforcing the safety of this weight-neutral approach. The program also demonstrated strong feasibility, achieving a recruitment rate of 82% (95/115) and a retention rate of 72% (69/95).

Accordingly, the current program consists of 3 structured sessions (outlined in Table 1). For this pilot study, a lower-intensity version of the program was implemented to assess its effectiveness and safety.

XSL•FO
RenderX

Table . Overview of the Green Apple program curriculum. The Green Apple program is a school-based, non–weight-centric educational intervention designed to promote metabolic health literacy among adolescents. The table outlines the content structure across 3 instructional units, including core topics and the rationale for each unit. All students received units 1 and 2; however, those in the enhanced intervention also received an additional unit 3 focused on the prevention of metabolic noncommunicable diseases.

Unit	Unit title	Topics	Rationale
Unit 1 (Green Apple)	Basics of nutrition	1. Carbohydrates 2. Proteins 3. Fats	To understand the relationship between macro and micronutrients, hunger and appetite, and convenience foods and whole foods.
Unit 2 (Green Apple)	Calories and energy	1. Energy in the body: how is energy produced, and why do we need it? 2. The relationship between food and energy 3. Where is energy stored in the body, and how is it released? 4. The relationship between physical activity and energy storage	To understand the relationship between macronutrients and energy sources, diet and exercise, and aerobic and weight-resistance exercises
Unit 3 (Enhanced Green Apple)	Metabolic chronic diseases management	1. Cholesterol: understanding its types and sources (beneficial and harmful), treatment with modern medicine, and lifestyle. 2. Diabetes: understanding types of diabetes and how modern medications and lifestyle influence blood sugar control. 3. How to help someone with MNCD[a].	To understand the mechanisms of stroke and diabetes and possible medical and lifestyle management plans.

[a]MNCD: metabolic noncommunicable disease.

Ethical Considerations

This study is part of a larger pilot initiative and was conducted in accordance with ethical standards approved by the Shaqra University Ethics Committee (approval ERC_SU_20230005). To ensure ethical treatment of participants, multiple procedures were implemented to safeguard their rights.

Before the intervention, written informed consent was obtained from the parents of all participating students. Consent was secured approximately 3 weeks before the program began and specifically covered both participation in the educational sessions and completion of the study surveys. The study's purpose, procedures, and voluntary nature were explained to students in an age-appropriate manner through both verbal presentations made by teachers and simple written materials provided to students and parents.

Participation was entirely voluntary. Students were clearly informed that choosing not to participate or choosing to skip any survey questions would not affect their academic grades. The intervention was delivered to all students within the classroom setting as part of the school health program; however, only students who provided informed consent and assent were included in the study analyses. Confidentiality and privacy were maintained by deidentifying all data before analysis. As a gesture of appreciation, all students received a small nonmonetary token (a keychain medley) upon completing the program surveys.

Measurements

Demographics

Baseline demographics included age, self-reported weight and height (if known), parents' education level, and monthly family income. Education level was categorized for both parents as either high (college degree and above) or low (high school degree and below). However, aside from age, rural students had limited knowledge of these details and could not complete this information, so it was excluded from the analysis.

Physical Activity

The Arab Teens Lifestyle Study (ATLS) includes a 22-item self-reported questionnaire designed to measure weekly physical activity by reporting the types and frequency of 11 specific physical activities (eg, walking, running, weight lifting, and housework) [34]. Then, the total energy expenditure in metabolic equivalent minutes per week (METs-min/wk) is calculated for high-vigorous physical activity of 6 METs and moderate-vigorous activity of 4 METs [34]. The ATLS also contains a 10-item self-reported eating habits questionnaire designed to measure 4 positive eating habits (eg, fruit and vegetable intake) and 6 negative eating habits (eg, sugary drink and fast food consumption) [34]. Scores are calculated based on the frequency of each habit per week. Total scores for positive eating habits (0 - 28) and negative eating habits (0 - 35) are generated based on frequency responses (eg, the number of days per wk). The ATLS has been specifically developed for

XSL•FO
RenderX

and validated among Arab adolescents, including those from Saudi Arabia [35].

Sedentary Behavior

The weekdays subscale of the self-reported Sedentary Behavior Questionnaire weekdays contains 9 items designed to measure sitting time during various sedentary activities (eg, watching television, playing computer or video games, listening to music, and reading books) during weekdays [36]. The average total sitting time is calculated based on 5 weekdays. Sitting ≥7 hours/day during weekdays is considered a highly sedentary behavior [37]. In our study, weekday sedentary behavior was measured at baseline and during follow-up. The Sedentary Behavior Questionnaire was originally developed for Western populations; however, its items have been culturally adapted and translated into Arabic, and it has been used among Saudi college students and adolescents with acceptable levels of reliability [33,36,38].

Exposure to Noncommunicable Diseases Health Education

Students' exposure to MNCD health education topics (eg, stroke and diabetes) was measured using a 4-item self-reported (yes or no) questionnaire developed by the researchers. A total score (0 - 4) was generated.

MNCD Knowledge

An 18-item true or false quiz was developed to assess students' knowledge of macronutrients and their role in MNCDs (eg, liver fat, vascular fat, diabetes, and hyperlipidemia). The items were aligned with the educational objectives of the Green Apple program. To ensure content validity and cultural relevance, the quiz underwent expert review by 2 nutrition professionals and a health education specialist. Items were reviewed for accuracy, relevance, and clarity in Arabic. The quiz was also piloted with a small group of students (n=12) to confirm age-appropriateness and understanding. Revisions were made to improve wording and comprehension. The quiz was intended to measure factual knowledge across multiple distinct subtopics (eg, liver fat, sugar, insulin, and vascular fat), rather than a single unidimensional construct. A total score ranging from 0 to 18 was calculated for each participant. The need for educational intervention was then evaluated using cutoff levels recommended in the guidelines of the Food and Agriculture Organization of the United Nations [39]. Using these guidelines, an urgent need for intervention was identified if 70% or fewer participants answered correctly. A considered need was indicated when 71% - 89% of participants answered correctly. No need for intervention was noted if 90% or more participants answered correctly. This scoring method enabled a targeted assessment of the gaps in students' knowledge of macronutrients and their role in MNCD prevention. MNCD knowledge was measured at baseline, postintervention, and during follow-up.

Body Image Discrepancy

Body image discrepancy was assessed using silhouettes representing 4 different body sizes [40] and a silhouette rating scale, which is a widely validated and reliable tool for assessing perceived and ideal body size among adolescents [41,42]. Silhouette scales are designed to be ethnically and culturally neutral, minimizing potential bias and ensuring applicability across diverse populations, including Saudi adolescents [43,44]. Students were asked to select the silhouette that best represented their ideal body size and the one they felt most closely resembled their current body size. The difference between the 2 selections (ideal vs current) was used to determine the degree of body image discrepancy. A negative score indicated a drive for thinness (desire for a smaller body size), while a positive score reflected a drive for fullness (desire for a larger body size) [43]. Body image discrepancy was measured at baseline, postintervention, and during follow-up.

Disordered Eating Symptoms

The Arabic version of the Sick-Control-One Stone-Fat-Food was used to assess disordered eating symptoms through a 5-item self-reported questionnaire with "yes or no" response options [45]. An answer of "yes"=1 and "no"=0, with a total score range from 0 - 5. Disordered eating symptoms were measured both at baseline and during follow-up. The Sick-Control-One stone-Fat-Food screening tool has been translated and validated in Arabic and used successfully with adolescents [45,46].

Statistical Approach

Descriptive statistics were first calculated for all study variables, including means and SDs for continuous variables and frequencies and percentages for categorical variables. Group differences at baseline were examined using independent samples *t* tests for continuous variables and chi-square tests for categorical variables. A series of linear mixed-effects models was used to examine the effects of time, intervention type, and school type on various outcome variables, while accounting for repeated measures across time points. The analyses were conducted using IBM SPSS Statistics (version 29) with restricted maximum likelihood, allowing for robust handling of the repeated measures design. For each model, estimated marginal means were calculated to explore pairwise comparisons between time points, intervention types, and school types. The Bonferroni correction was applied to adjust for multiple comparisons. The goodness-of-fit for each model was evaluated using the Akaike information criterion and Schwarz's Bayesian criterion. In total, 4 outcome variables were examined: knowledge, body image discrepancy, disordered eating symptoms, and sedentary behavior. Each model included time (baseline, postintervention, or follow-up), intervention type (Green Apple-MNCD or Green Apple), and school type (urban or rural) as categorical fixed effect variables. Repeated measurements were modeled with time as the repeated factor and student ID as the subject variable to account for the correlation between repeated measurements from the same individual. An unstructured covariance matrix was used to account for variability at different time points. Pairwise comparisons were conducted for both the 2-way interaction (time×intervention type) and the 3-way interaction (time×intervention type×school type). In addition, for the knowledge model, exposure to MNCD health education was included as a continuous covariate to control baseline health literacy. However, for the body image model, knowledge was included as a continuous predictor to assess its effect on body image. For the disordered eating symptoms and sedentary

XSL•FO
RenderX

behavior models, there were no additional covariates. Statistical significance was determined at $P<.05$.

Sample Size

The sample size for this study was calculated using G*Power 3.1.9.7 software (Heinrich-Heine-Universität Düsseldorf), assuming a repeated measures ANOVA design to approximate the linear mixed model framework. A medium effect size (f=0.25), a significance level (α=.05), and a power (1−β=.80) were used. Although our previous pilot evaluation of the same program demonstrated a large effect size (partial η^2=0.27, equivalent to f≈0.61), we selected a more conservative medium effect size (f=0.25) for the current sample size estimation [33]. This decision was based on both methodological caution and practical considerations, as we anticipated a relatively small sample due to school and class availability constraints in rural areas. The design included 2 groups and 3 time points (pre, post, and follow-up), with an assumed correlation of 0.5 among repeated measures. The analysis indicated that a total sample size of 86 participants was required to achieve sufficient power. However, when adjusted for an expected attrition rate of 20%, the required sample size increased to 108 participants.

Missing Data

An intention-to-treat approach was applied to ensure that all participants were analyzed in their original groups, even if they had missing data at later stages. Missing data were handled using restricted maximum likelihood, which is appropriate for data assumed to be missing at random in linear mixed models [47]. This method enables the use of all available data without requiring imputation or exclusion of cases with missing values, resulting in robust parameter estimation. No imputation techniques were applied, as linear mixed models inherently accommodate incomplete data in repeated measures by using the available observations for each participant [47]. Model fit was evaluated using the Akaike information criterion, and likelihood ratio tests were performed to compare nested models. The normality of residuals and homoscedasticity was checked graphically using residual plots, and no major violations of assumptions were observed.

Data Collection

Surveys were administered in paper-based format and completed during regular class periods under the supervision of research assistants (registered nurses), who provided clarification when necessary. Baseline surveys required approximately 45 minutes to complete, while postintervention and follow-up surveys required 15 - 20 minutes. Upon collection, all surveys were collected immediately and reviewed for completeness and then entered manually into a secure electronic database (Microsoft Office Forms) by the research assistants. To minimize entry errors, data were double-checked, and any discrepancies were resolved through cross-checking.

Results

Sample Baseline Characteristics

Demographics by School Type

A total of 107 students agreed to participate in the study. Furthermore, 2 cases from the urban school were excluded due to missing identification, resulting in 105 students included in the final analysis (refer to Multimedia Appendix 1 for details on data completeness). Overall, 7% of data were missing, with the majority occurring at the follow-up stage (9%). Little's Missing Completely at Random test indicated that the data were missing completely at random (χ^2_{111}=124.7; P=.18), suggesting that the missingness was unlikely to introduce systematic bias. All students contributed data at a minimum of two time points.

The mean age across all participants was 16.42 (SD 0.66) years (Table 2). A significant difference in age between the intervention groups was found ($P<.001$). Combining both urban and rural schools, the overall mean age for the enhanced intervention (Green Apple+MNCD) group was 15.97 (SD 0.41) years; for the intervention (Green Apple) group, the mean age was 17.00 (SD 0.42) years. For additional baseline characteristics by intervention type, refer to Multimedia Appendix 2.

XSL•FO
RenderX

Table . Sample sizes and mean ages of students across intervention and school types. The study includes participants from both urban and rural schools, with 2 types of interventions: enhanced intervention (Green Apple+metabolic noncommunicable disease) and intervention (Green Apple).

School type and types of intervention	Participants, n	Age (years), mean (SD)
Urban school		
Green Apple+MNCD[a]	25	15.92 (0.49)
Green Apple	15	17.13 (0.35)
Total	40	16.38 (0.74)
Rural school		
Green Apple+MNCD	34	16.10 (0.35)
Green Apple	31	16.94 (0.44)
Total	65	16.45 (0.61)
Total		
Green Apple+MNCD	59	15.97 (0.41)
Green Apple	46	17.10 (0.42)
Total	105	16.42 (0.66)

[a]MNCD: metabolic noncommunicable disease.

Physical Activity and Sedentary Behavior

Table 3 presents a comparison of urban and rural schools regarding various physical activity behaviors, exposure to MNCD health education, and the need for educational interventions. Urban students reported a higher mean sedentary time than rural students; however, this difference was not statistically significant (*P*=.09). Urban students exhibited significantly higher total energy expenditure in METs-minutes/week than rural students (*P*=.03). This was primarily due to a significantly higher energy expenditure from vigorous physical activity (*P*=.009). Although urban students reported higher energy expenditure from moderate-vigorous physical activity, this difference was not statistically significant (*P*=.17).

XSL•FO
RenderX

Table . Baseline descriptive statistics by school type (urban vs rural): physical activity, exposure to metabolic noncommunicable diseases, health education, knowledge, body image, eating disorder symptoms, and perceived need for intervention. *P* values reflect between-group comparisons using independent sample *t* tests or chi-square tests where appropriate.

Variable	Urban school	Rural school	*P* value
Physical activity behaviors, mean (SD)			
Sedentary behavior (h/d)	7.85 (3.17)	6.96 (3.68)	.09
Total-intensity physical activity (METs[a]-min/wk)	3092.62 (3025.67)	2127.11 (2217.04)	.03[b]
Vigorous-intensity physical activity (METs-min/wk)	1370.75 (2042.03)	667.11 (961.12)	.009[b]
Moderate-intensity physical activity (METs-min/wk)	1592.33 (1433.34)	1353.43 (1166.18)	.17
Exposure to MNCDs[c] health education	3.68 (0.53)	3.38 (0.93)	.02[b]
MNCDs knowledge, mean (SD)	11.78 (2.51)	10.14 (2.34)	<.001[b]
Body image discrepancy, mean (SD)	−0.769 (0.8)	−0.23 (0.98)	.002[b]
Eating disorder symptoms, mean (SD)	1.93 (1.16)	1.37 (1.04)	.006[b]
Need of intervention assessment, n (%)			
Urgent need	21 (52.5)	57 (87.7)	<.001[b]
Considered need	19 (47.5)	8 (12.3)	—[d]
No need	0 (0)	0 (0)	—

[a]MET: metabolic equivalent.

[b]Indicates statistically significant differences (*P*<.05).

[c]MNCD: metabolic noncommunicable disease.

[d]No measurement.

Exposure to MNCD Health Education

Urban students reported significantly higher exposure to MNCD health education than rural participants, with a significant *P* value (*P*=.02). Furthermore, urban students reported significantly higher MNCD knowledge than rural participants (*P*<.001) (Table 3).

Need for Intervention Assessment

A notable disparity was observed between urban and rural schools regarding the perceived need for an intervention (Table 3). In urban schools, 53% (21/40) of participants reported an urgent need for intervention, while 48% (19/40) considered the need but did not find it urgent. In rural schools, 88% (57/65) of participants reported an urgent need, with only 12% (8/65) considering the need but not finding it urgent. This difference was significant (*P*<.001).

Body Image Discrepancy and Disordered Eating Symptoms

Urban students demonstrated a significantly greater drive for thinness (*P*=.002) and more disordered eating symptoms (*P*=.006) than rural students (Table 3). This suggests that urban students may be more vulnerable to body image dissatisfaction and disordered eating symptoms, highlighting a potential need for targeted interventions.

Green Apple Intervention

Program Effectiveness on Knowledge Level

At all 3 time points (baseline, postintervention, and follow-up), there were no statistically significant differences between the enhanced intervention (Green Apple+MNCD) group and the original intervention (Green Apple) group in knowledge about MNCDs (all *P* values >.05; Table 4). This lack of significance may reflect the substantial overlap in core educational content between the 2 interventions, particularly regarding macronutrient understanding and behavior change strategies.

XSL•FO
RenderX

Table . Within-group school differences: changes from baseline to postintervention and to follow-up. Estimates are derived from linear mixed-effects models controlling for time and school type. Values represent adjusted mean differences with 95% CI.

Change	Change from baseline to post		Change from baseline to follow-up	
	Adjusted mean difference (95% CI)	*P* value	Adjusted mean difference (95% CI)	*P* value
On knowledge about NCDs[a] score				
Enhanced intervention (Green Apple+MNCDs[b])				
Urban school	2.70 (1.21 to 4.19)	<.001[c]	1.83 (0.27 to 3.39)	.02[c]
Rural school	1.00 (−0.20 to 2.20)	.13	1.48 (0.09 to 2.88)	.03[c]
All schools	1.85 (0.89 to 2.81)	<.001[c]	1.66 (0.61 to 2.70)	<.001[c]
Green Apple				
Urban school	1.86 (−0.08 to 3.81)	.06	1.22 (−0.82 to 3.25)	.44
Rural school	0.67 (−0.61 to 1.96)	.61	1.03 (−0.204 to 2.39)	.20
All schools	1.27 (0.10 to 2.43)	.02[c]	1.13 (−0.10 to 2.35)	.08
On body image discrepancy				
Enhanced Intervention (Green Apple+MNCDs)				
Urban school	0.020 (−0.59 to 0.55)	>.99	0.09 (−0.43 to 0.60)	>.99
Rural school	−0.12 (−0.59 to 0.35)	>.99	0.04 (−0.41 to 0.49)	>.99
All schools	−0.05 (−0.42 to 0.33)	>.99	0.06 (−0.29 to 0.41)	>.99
Green Apple				
Urban school	−0.01 (−0.74 to 0.72)	>.99	0.42 (−0.24 to 1.08)	.36
Rural school	−0.08 (−0.59 to 0.43)	>.99	0.19 (−0.36 to 0.56)	>.99
All schools	−0.05 (−0.49 to 0.40)	>.99	−0.26 (−0.67 to 0.14)	.35
On DE[d] symptoms				
Enhanced intervention (Green Apple+MNCDs)				
Urban school	—[e]	—	−0.94 (−1.61 to −0.26)	.007[c]
Rural school	—	—	0.09 (−0.53 to 0.69)	.78
All schools	—	—	−0.43 (−0.88 to 0.03)	.06
Green Apple				
Urban school	—	—	0.58 (−0.30 to 1.46)	.19
Rural school	—	—	−0.07 (−0.51 to 0.65)	.80
All schools	—	—	0.25 (−0.27 to 0.78)	.34
On sedentary behavior				
Enhanced intervention (Green Apple+MNCDs)				
Urban school	—	—	−1.01 (−3.94 to 1.93)	.49
Rural school	—	—	−2.05 (−4.69 to −0.59)	.12
All schools	—	—	−1.53 (−3.50 to 0.45)	.12
Green Apple				
Urban school	—	—	−1.67 (−5.59 to 2.24)	.39
Rural school	—	—	−3.12 (−5.67 to −0.57)	.02[c]

XSL•FO
RenderX

Change	Change from baseline to post		Change from baseline to follow-up	
	Adjusted mean difference (95% CI)	*P* value	Adjusted mean difference (95% CI)	*P* value
All schools	—	—	−2.39 (−4.73 to −0.06)	.04[c]

[a]NCD: noncommunicable disease.

[b]MNCD: metabolic noncommunicable disease.

[c]Indicates statistically significant differences (*P*<.05).

[d]DE: disordered eating.

[e]No measurement.

Although the enhanced intervention (Green Apple+MNCD) group received additional disease-specific content, the intervention intensity (1 extra session) may not have been sufficient to produce a measurable advantage, especially given the short follow-up period and modest sample size (refer to Multimedia Appendix 3 for details on types of units included within each intervention).

The direction of the difference at follow-up favored the enhanced intervention (Green Apple+MNCD) group (mean difference 0.73, SE 0.49, 95% CI –0.24 to 1.71), suggesting a possible trend that may warrant further investigation in larger or longer-duration trials. The enhanced intervention (Green Apple+MNCD) led to significant improvements in knowledge about MNCDs across all schools, with an adjusted mean difference from baseline to postintervention of 1.85 (95% CI 0.89 - 2.81; *P*<.001) and from baseline to follow-up of 1.66 (95% CI 0.61 - 2.70; *P*<.001) (refer to Table 4 for within-group school differences in knowledge scores). In addition, when effectiveness was examined based on school type, it was found that the enhanced intervention (Green Apple+MNCD) led to significant improvements in urban schools from baseline to both postintervention with an adjusted mean difference of 2.70 (95% CI 1.21 - 4.19; *P*<.001) and remained significantly higher at follow-up with an adjusted mean difference of 1.83 (95% CI 0.27 - 3.39; *P*=.02). In rural schools, the enhanced intervention (Green Apple+MNCD) showed no significant improvement at postintervention with an adjusted mean difference of 1.00 (95% CI –0.20 to 2.20; *P*=.13), but a delayed significant improvement was observed at follow-up with an adjusted mean difference of 1.48 (95% CI 0.09 to 2.88; *P*=.03).

On the other hand, the intervention (Green Apple) group showed a modest overall improvement from baseline to postintervention across all schools, with an adjusted mean difference of 1.27 (95% CI 0.10 - 2.43; *P*=.02). However, this was not maintained at follow-up, where the improvement was not statistically significant, with an adjusted mean difference of 1.13 (95% CI –0.10 to 2.35; *P*=.08). Furthermore, when this effectiveness was examined based on school type, it was found that in urban schools, the intervention (Green Apple) group showed marginally nonsignificant improvement at postintervention, with an adjusted mean difference of 1.86 (95% CI –0.08 to 3.81; *P*=.06), and nonsignificant improvement at follow-up, with an adjusted mean difference of 1.22 (95% CI –0.82 to 3.25; *P*=.44). Similarly, in rural schools, the intervention (Green Apple) group showed no significant improvement at postintervention (adjusted mean difference 0.67, 95% CI –0.61 to 1.96; *P*=.61) or follow-up (adjusted mean difference 1.03, 95% CI –0.204 to 2.39; *P*=.20), further supporting the conclusion of limited effectiveness.

Program Safety for Body Image Discrepancy and Disordered Eating Symptoms

A key aim of this study was to determine whether the program, including both the intervention (Green Apple) and the enhanced intervention (Green Apple+MNCD), maintained a neutral effect on body image discrepancy and disordered eating symptoms, as indicated by a lack of significant changes. Results across various school settings, both urban and rural, provide insight into whether these interventions unintentionally influence drive for thinness or disordered eating behaviors, which would indicate a safety concern. First, at baseline, there was no significant difference in body image discrepancy between the enhanced intervention (Green Apple+MNCD) group and the intervention (Green Apple) group (mean difference –0.05, SE 0.19, 95% CI –0.42 to 0.33; *P*=.80). This similarity at baseline confirms that the groups were comparable in this outcome before the intervention. Postintervention, the difference remained nonsignificant (mean difference –0.04, 95% CI –0.41 to 0.32; *P*=.81). However, at follow-up, a statistically significant difference was observed in favor of the enhanced intervention (Green Apple+MNCD) group (mean difference –0.37, SE 0.19, 95% CI –0.73 to –0.001; *P*=.05), indicating a greater reduction in body image discrepancy compared with the intervention (Green Apple) group.

Although the effect size was small, this finding suggests a potential delayed benefit of integrating MNCD-focused content within a weight-neutral framework. Across both intervention types and school settings, there was a general trend toward a reduction in body image discrepancy; however, none of the observed changes from baseline to follow-up reached statistical significance. For instance, in the intervention (Green Apple) group, the overall mean difference from baseline to follow-up across all schools was –0.26 (95% CI –0.67 to 0.14; *P*=.35), suggesting a modest improvement in body image perception that did not reach statistical significance. Similarly, the enhanced intervention (Green Apple+MNCD) group showed a slight increase in body image discrepancy across all schools, with a mean difference of 0.06 (95% CI –0.29 to 0.41; *P*>.99). When the data were disaggregated by school type, urban and rural schools in both interventions reflected similar nonsignificant trends. For example, in urban schools, the intervention (Green Apple) group showed a mean change of 0.42 (95% CI –0.24 to 1.08; *P*=.36), while the enhanced intervention (Green Apple+MNCD) group showed a mean change of 0.09 (95% CI

XSL•FO
RenderX

–0.43 to 0.60; *P*>.99). These results suggest a generally neutral effect on body image discrepancy across both interventions, although the trends, particularly the reduction observed in the intervention (Green Apple) group, indicate potential in reducing the drive for thinness without causing psychological harm (refer to Table 5 for between-group differences and Table 4 for within-group differences in body image discrepancy).

Table . Adjusted between-group mean differences (enhanced intervention: Green Apple+MNCD vs intervention: Green Apple) at baseline, postintervention, and follow-up for key outcomes. Estimates are derived from linear mixed-effects models controlling for time and school type. Values represent adjusted mean differences with SEs and 95% CIs.

Variables	Baseline			Postintervention			Follow-up		
	Mean (SE)	95% CI	*P* value	Mean (SE)	95% CI	*P* value	Mean (SE)	95% CI	*P* value
Knowledge about MNCDs[a]	0.20 (0.46)	−0.71 to 1.12	.65	0.79 (0.45)	−0.10 to 1.7	.83	0.73 (0.49)	−0.24 to 1.71	.13
Body image discrepancy	−0.05 (0.19)	−0.42 to 0.33	.80	−0.04 (0.19)	−0.41 to 0.32	.81	−0.37 (0.19)	−0.73 to 0.001[b]	.05[b]
DE[c] symptoms	0.30 (0.23)	−0.15 to 0.75	.18	—[d]	—	—	−0.38 (0.27)	−0.92 to 0.16	.16
Sedentary behavior	−0.07 (0.99)	−2.04 to 1.90	.94	—	—	—	0.80 (1.09)	−1.36 to 2.96	.46

[a]MNCDs: metabolic noncommunicable diseases.

[b]Indicate statistically significant differences (*P*<.05).

[c]DE: disordered eating.

[d]No measurement.

Despite the absence of statistically significant change, the observed decrease in discrepancy scores may still hold clinical relevance, as even modest reductions in the drive for thinness could reflect a shift toward healthier body image perceptions. The lack of significance may be attributed to the relatively short intervention and follow-up duration, the complexity of modifying entrenched body image attitudes, or limited statistical power to detect small effects. Second, both the enhanced intervention (Green Apple+MNCD) and the intervention (Green Apple) demonstrated no significant increases in disordered eating symptoms across various settings, supporting the safety of these weight-neutral programs (refer to Table 5 for between-group differences and Table 4 for within-group differences in disordered eating symptoms). At baseline, disordered eating symptoms were slightly higher in the enhanced intervention (Green Apple+MNCD) group than in the intervention (Green Apple) group across all schools, with a nonsignificant mean difference of 0.30 (95% CI −0.15 to 0.75; *P*=.18), suggesting comparable starting levels between the groups (Table 5). At follow-up, the enhanced intervention (Green Apple+MNCD) group showed a slight reduction in mean disordered eating symptoms, while the intervention (Green Apple) group showed a slight increase. However, neither of these changes reached statistical significance (*P*=.17 for between-group comparison at follow-up), supporting the safety of the interventions in not exacerbating disordered eating symptoms.

Across all schools, the enhanced intervention (Green Apple+MNCD) group showed a slight, nonsignificant reduction in disordered eating symptoms from baseline to follow-up (mean difference −0.43, 95% CI −0.88 to 0.03; *P*=.06). In urban schools, the enhanced intervention (Green Apple+MNCD) group demonstrated a significant reduction in disordered eating symptoms (mean difference −0.94, 95% CI −1.61 to −0.26; *P*=.007), suggesting a positive safety profile. In rural schools, the enhanced intervention (Green Apple+MNCD) group showed a nonsignificant change in symptoms from baseline to follow-up (mean difference 0.09, 95% CI −0.53 to 0.69; *P*=.78), possibly reflecting contextual differences in program delivery or participant characteristics. Similarly, the intervention (Green Apple) group in both urban and rural schools showed no statistically significant changes in symptoms, reinforcing the safety of the program's weight-neutral design concerning disordered eating symptoms. However, it is important to note that both programs were weight-neutral; therefore, a significant difference between the 2 intervention arms was not necessarily expected. The significant reduction observed in urban schools within the enhanced intervention (Green Apple+MNCD) group may be partially explained by the higher baseline mean of disordered eating symptoms in this subgroup compared with others. This suggests that the intervention may have been particularly beneficial for students with elevated symptoms at baseline, while producing few or no significant changes in groups with lower levels of disordered eating symptoms.

Program Safety on Sedentary Behavior

This study aimed to assess the impact of the intervention (Green Apple) and enhanced intervention (Green Apple+MNCD) on sedentary behavior, defined as hours spent sitting per day during weekdays, with 7 or more hours defined as sedentary. A decrease in hours would suggest a positive effect on reducing sedentary behavior, supporting the interventions' safety to not exacerbate sedentary tendencies. Overall, there was a trend toward decreased sedentary behavior across all schools in both intervention groups, with the enhanced intervention (Green Apple+MNCD) and intervention (Green Apple) groups showing

XSL•FO
RenderX

reductions in mean sedentary hours from baseline to follow-up (refer to Table 5 for between-group differences and Table 4 for within-group differences in sedentary behavior). Across all schools, the enhanced intervention (Green Apple+MNCD*)* group showed a decrease in sedentary behavior from baseline (mean 10.22 h) to follow-up (mean 8.69 h), though this change was not statistically significant (mean difference −1.53, 95% CI −3.50 to 0.45; *P*=.12). In the intervention (Green Apple) group, sedentary behavior declined significantly across all schools (mean difference −2.39, 95% CI −4.73 to −0.06; *P*=.04), suggesting a positive impact.

When the data were examined by school type, both the enhanced intervention (Green Apple+MNCD) and intervention (Green Apple) groups in urban schools showed reductions in sedentary behavior from baseline to follow-up, although these changes were not statistically significant (*P*=.49 and *P*=.39, respectively). In rural schools, the intervention (Green Apple) group exhibited a statistically significant reduction in sedentary behavior (mean difference −3.12, 95% CI −5.67 to −0.57; *P*=.02). Meanwhile, the enhanced intervention (Green Apple+MNCD) group showed a nonsignificant decrease in sedentary behavior in rural schools (mean difference −1.05, 95% CI −4.69 to 2.59; *P*=.12), which supports the program's safety. The nonsignificant difference between the 2 intervention groups regarding sedentary behavior may be attributed to the overlap in content related to calories and energy. Both interventions included the same unit discussing how energy is produced, where it is stored in the body, how it is released, and how physical activity and energy balance are related. The only distinction is that the enhanced intervention (Green Apple+MNCD) additionally linked this unit to the management of MNCDs.

Discussion

Principal Findings

This study found that the Green Apple program, a non–weight-centric school-based intervention, significantly improved students' knowledge of MNCDs, reduced sedentary behavior — especially among rural students — and showed promising trends in improving body image discrepancy and disordered eating symptoms without causing harm. These findings support the effectiveness and safety of implementing culturally tailored, weight-neutral interventions in adolescent populations.

To our knowledge, this implementation trial is one of the first Arabic-language, school-based programs to evaluate a weight-neutral approach for improving understanding of MNCDs — particularly metabolic conditions like diabetes, stroke, and hyperlipidemia — while focusing on visceral fat and lifestyle changes rather than weight loss among rural and urban female adolescents. Developed using a theoretically guided process and formative research, the intervention prioritized health-promoting behaviors without reinforcing weight stigma.

Comparison With Previous Work

The Green Apple program's effectiveness in improving MNCD-related knowledge without exacerbating sedentary behavior tendencies can be attributed to several targeted components. First, the curriculum emphasized interactive and culturally tailored lessons that linked macronutrient intake with specific metabolic health outcomes, such as liver health and vascular fat accumulation — making abstract biomedical concepts more relevant and accessible to adolescents. Second, rather than prescribing dietary changes or promoting physical activity directly, the program employed a health literacy approach that fostered critical thinking and informed decision-making. For example, students were taught to distinguish between aerobic and resistance exercises and to identify how each affects energy metabolism. This empowered students to re-evaluate their daily habits, which may explain the observed decrease in sedentary behavior without an explicit call to increase activity. The use of relatable examples, age-appropriate language, and visual aids (eg, infographics and simplified diagrams) also played an important role in reinforcing key messages and supporting knowledge retention. The Green Apple program engaged students not only as passive recipients of information but as active participants in understanding their own health behaviors, which may have been instrumental in achieving both cognitive (knowledge) and behavioral (sedentary time) changes.

The Green Apple program was developed with a foundation in the HBM, SCT, and the TTM, aiming to enhance metabolic health literacy, self-efficacy, and staged behavior change among adolescents. Beyond these, Green Apple shares theoretical intersections with frameworks frequently used in international weight-neutral interventions. For instance, elements of self-determination theory are reflected in the program's emphasis on autonomy (students are encouraged to make informed health decisions rather than passively follow directives), building competence (through skill-based, practical lessons), and fostering relatedness (via collaborative classroom activities) [48]. Similarly, social learning theory is embodied through its use of observational and peer-supported learning, including interactive discussions and role-modeling by program facilitators [49]. While acceptance and commitment therapy was not an explicit design framework, the program's avoidance of weight-centric language and focus on values-driven, health-promoting behaviors may support the psychological flexibility and acceptance central to acceptance and commitment therapy–based interventions [50]. By integrating these diverse theoretical perspectives, Green Apple addresses multiple drivers of adolescent health behavior in a culturally sensitive manner.

The Green Apple program's effectiveness in enhancing students' understanding of MNCDs aligns with previous research highlighting the importance of health education in obesity prevention [16,51]. By focusing on the role of liver fatty acids and vascular fat in metabolic conditions rather than on weight loss, the program shifts the conversation away from potentially stigmatizing weight-centric approaches. This approach supports recent calls to reshape the global obesity narrative to recognize and reduce weight stigma [1,10,52]. The program's success in increasing knowledge about MNCDs suggests that high school students can comprehend topics like metabolic diseases and their connection to macronutrients and body energy. This

XSL•FO
RenderX

education-focused intervention demonstrates the potential of such strategies in promoting health literacy among adolescents.

Recent years have seen a global shift toward school-based health interventions that adopt weight-neutral approaches, emphasizing overall well-being, health behaviors, and body acceptance rather than weight loss or BMI reduction [53,54]. Innovative programs like Good Food for All (Australia) and Imagine HEALTH (United States) incorporate experiential activities, intuitive eating, and guided imagery to foster positive relationships with food and movement, improving psychosocial health and lifestyle behaviors without targeting weight loss [54,55]. Mindful eating curricula, such as "Eat My ABCs," further extend these principles by teaching self-regulation and satiety awareness, leading to healthier eating behaviors in children [54]. While these programs and other international interventions share key characteristics with the weight-neutral, stigma-reducing strategies now associated with the mind-body modalities of the HAES paradigm [53,54], they are not always formally aligned with or derived from HAES.

The Green Apple program shares similarities with the HAES movement [54] in that both emphasize promoting health without focusing solely on weight, avoiding calorie counting and restrictive dieting, and encouraging a healthy relationship with food. However, there are important distinctions between the 2. The Green Apple program was specifically designed for school settings and is driven by the goal to increase metabolic health literacy among adolescents, particularly in relation to macronutrients, energy metabolism, and lifestyle-related noncommunicable diseases (eg, diabetes and hypertension). In contrast, HAES is a broader, philosophically driven approach that promotes body inclusivity, intuitive eating, and joyful movement across all populations, regardless of age, gender, or health status [53,54]. While HAES has been praised for reducing weight stigma and supporting psychological well-being, some critics argue that its generalizability to targeted disease prevention (eg, metabolic syndrome) may be limited [56,57]. On the other hand, Green Apple's structured and curriculum-based model offers clear educational content aligned with disease prevention but may require more rigorous implementation and contextual adaptation.

Furthermore, the Green Apple program may be more culturally adaptable in non-Western contexts, such as Saudi Arabia, where public health messages around metabolic disease are often aligned with national health priorities [14,15]. In contrast, HAES's strong focus on body autonomy and its rejection of medical weight loss may face cultural or institutional resistance in certain settings. In summary, both programs focus on health promotion beyond weight loss, but the Green Apple program — specifically, the enhanced intervention (Green Apple+MNCD) used in our study— uniquely targets metabolic health literacy and prevention, making it more tailored for specific outcomes related to metabolic disease management in adolescent females. In a previous pilot study, the Green Apple+MCND combination showed greater pre-post effectiveness on knowledge than the Green Apple program alone among male and female urban adolescents [33]. Thus, the findings from this study contribute to the literature by highlighting the effectiveness of this program among rural students and its sustainability at follow-up.

Future studies should explore the program's effectiveness among rural male adolescents and its sustainability among urban male adolescents. Notably, rural students in this study demonstrated a delayed improvement in nutrition knowledge, with no significant gains immediately following the intervention but a marked increase at follow-up. This pattern may indicate a need for more time to absorb and consolidate new information, particularly among populations with lower baseline health literacy [22,58]. Overall, the Green Apple program demonstrated its universal acceptability among rural female adolescents and its long-term effectiveness during follow-up among both rural and urban students.

The trend reduction in sedentary behavior observed among participants is particularly noteworthy, given the high prevalence of sedentary lifestyles among Saudi youth [59]. At baseline, all schools exhibited sedentary behavior (over 7 h of sitting daily). By follow-up, sedentary time decreased by 1 - 3 hours per weekday, despite the program not explicitly instructing students to reduce sitting time or increase physical activity. Several mechanisms may explain this outcome. First, the program emphasized the physiological relationship between macronutrients and energy usage, highlighting how different types of physical activity (eg, aerobic vs resistance training) impact health. This knowledge-based approach may have fostered greater internal motivation and self-efficacy to engage in movement, leading students to make conscious efforts to reduce sedentary time. Second, increased awareness of energy balance and muscle health could have encouraged students to re-evaluate their daily habits, choosing more active routines in response to improved health literacy. Similar findings have been observed in other weight-neutral interventions [54]. For example, a pilot RCT involving Latino adolescents with BMIs >95th percentile found that those who attended 12 weekly sessions of a non–weight-centric lifestyle education program called H.E.A.L.T.H. that incorporated guided imagery experienced significant reductions in leisure sedentary behavior (P<.05) compared with adolescents (n=14) who attended the same program with digital storytelling sessions [55]. Future studies should further explore these mechanisms by assessing not only the types and duration of physical activity adopted but also changes in students' beliefs, confidence, and motivations regarding movement. To sum up, these findings support the potential of non–weight-centric school-based interventions to meaningfully reduce sedentary time through education and empowerment, rather than prescriptive behavioral mandates.

Finally, the promising trends in decreasing body image discrepancy and disordered eating symptoms in this study are particularly noteworthy, given the potential negative consequences of weight stigma and body dissatisfaction on adolescent mental health [2,60]. Although these changes were not statistically significant, they may still be clinically meaningful and suggest that weight-neutral approaches like the Green Apple program can support the development of positive body image and healthy eating behaviors without exacerbating weight concerns. These findings highlight the potential of such interventions to contribute to the prevention of disordered eating

XSL•FO
RenderX

symptoms, even in the absence of immediate measurable effects. This aligns with recent research supporting the efficacy of weight-inclusive interventions in enhancing psychological well-being and reducing disordered eating behaviors [6,54,60]. For example, a meta-analysis found nonsignificant trends toward improvements in drive for thinness and body dissatisfaction with weight-neutral interventions [6]. However, consistent with our study, a statistically significant improvement in disordered eating symptoms was observed within one of the weight-neutral intervention groups [6].

This study's results also highlight the importance of tailoring interventions to specific cultural contexts. The observed differences between urban and rural schools highlight the need for tailored interventions based on regional disparities. Urban students reported higher levels of knowledge and exposure to health education, along with higher body image discrepancy and disordered eating symptoms, while rural students exhibited a greater need to increase their knowledge about MNCDs. Body image is strongly shaped by social and cultural influences, and this program offers a valuable starting point for future interventions that aim to include more tailored strategies for addressing body image concerns, particularly in rural settings. This study observed an average discrepancy of less than 1 body size among students, which is consistent with previous findings suggesting that body image discrepancy tends to be low among rural adolescents [61,62]. As such, future iterations of the Green Apple program may benefit from further customization to address the specific needs of rural populations, ensuring that interventions are both accessible and effective across diverse settings. Furthermore, the use of silhouette scales, while widely accepted for body image assessment, may be limited by participants' memory effects. Damasceno et al [63] found that participants often selected the same or very similar silhouettes within a 15-day interval, suggesting that repeated exposure to the same silhouettes over short periods can artificially inflate reliability and reduce sensitivity to actual change. This may have contributed to the lack of observed changes in body image over time.

Overall, the results in both settings demonstrate the potential for adapting weight-neutral approaches to non-Western cultures. This is particularly relevant given the limited research on obesity prevention in Saudi Arabia and the need for culturally appropriate interventions. The program's effectiveness in both urban and rural schools further underscores its adaptability and potential for widespread implementation.

However, the observed differences between urban and rural schools warrant further exploration. Urban students generally had higher baseline levels of health literacy, greater previous exposure to health education, and better access to digital and extracurricular resources, which may have facilitated quicker knowledge gains and behavioral changes [58,64]. In contrast, rural students demonstrated a delayed improvement in nutrition knowledge, with gains becoming evident only at the follow-up stage. This delay may reflect disparities in educational resources, differences in school infrastructure, or cultural variations in how health information is received and processed [58,64]. These findings emphasize the need to tailor the program's delivery with reinforcement strategies to better support students in underresourced rural settings and reduce health literacy gaps between regions [58,64].

Public Health Policy Recommendations

The findings of this study suggest a significant need for policy reform in school-based health interventions, particularly in how obesity prevention is addressed in Saudi Arabia.

First, adopt non–weight-centric frameworks in school health programs. Current school health initiatives should transition from weight-focused models to non–weight-centric approaches that emphasize metabolic health. Rather than using BMI as the primary indicator of health, schools can implement annual wellness screenings that assess cardiovascular fitness, physical activity habits, and dietary quality [1,65-67]. For instance, in collaboration with schools, community health nurses, physical education staff, and local primary health care centers can adopt metrics such as heart rate recovery, physical endurance, and nutrition intake to provide a more holistic and less stigmatizing view of adolescent health.

Second, revise educational content to use stigma-free, health-promoting language. Health education policies should be revised to eliminate language that links health behaviors explicitly to weight loss. Instead, curricula should emphasize building healthy habits, such as engaging in enjoyable physical activity, eating diverse whole foods, and understanding how nutrition supports organ health [1,3]. For example, Ministry of Education health textbooks should be updated, and teacher development workshops on inclusive language should be delivered. These workshops can include scenario-based training and feedback sessions to help educators identify and eliminate implicit weight bias in their teaching [66,68].

Third, tailor health interventions to address regional disparities. As this study indicates, rural areas often lack access to tailored health education programs. To bridge this gap, mobile regional health directorates can deploy school health units to deliver interactive workshops and preventive screenings in underserved schools. In addition, rural teachers can be trained as "health ambassadors" in collaboration with local universities, ensuring consistent delivery of health messages that are culturally and contextually appropriate. Pilot programs in targeted rural regions can serve as models for broader scale-up [10,69,70].

Fourth, support long-term sustainability through policy and resource allocation. Health policies should ensure that resources are allocated for ongoing support, including teacher training and curriculum development, to maintain the positive outcomes seen in short-term interventions [10,66,68].

Fifth, encourage research and innovation in weight-neutral health promotion. Government agencies and research institutions should allocate additional resources to support studies that explore weight-neutral approaches to health promotion and obesity prevention [10,66]. National grants can prioritize school-based interventions that focus on improving metabolic indicators without relying on weight loss as an outcome. In addition, research institutions can be encouraged to collaborate with school systems to cocreate evidence-based, student-centered health programs. Finally, evaluating the Green Apple program's adaptability to other adolescent health issues,

XSL•FO
RenderX

such as iron deficiency anemia, malnutrition, or diabetes management, could broaden its impact and support more comprehensive school-based health education efforts, especially given the limited availability of evidence-based interventions targeting metabolic health in the current Saudi literature [71].

Limitations and Recommendations

While this study presents promising results, several limitations must be acknowledged. First, the relatively short follow-up period limits our ability to assess the long-term impact and sustainability of the intervention. Some of the nonsignificant findings — particularly in outcomes such as body image discrepancy and disordered eating symptoms — may be attributed to the brief duration of the intervention, which may not have been sufficient to elicit measurable changes in complex psychosocial constructs. Future research should incorporate longer follow-up periods to better capture delayed or long-term effects. Second, the modest sample size may have reduced the statistical power to detect small but meaningful effects, especially in subgroup analyses and interaction terms. Finally, the similarity in content between the 2 intervention arms — both of which adhered to a non–weight-centric framework — may have reduced the likelihood of observing substantial between-group differences, as both emphasized promoting healthy behaviors without explicitly targeting weight loss or body image modification. In addition, since this study focused only on female students in Saudi Arabia, its findings may not be generalizable to other populations. Further research is needed to test similar interventions across diverse cultural settings and among different gender groups.

Despite these limitations, the study provides a significant contribution to adolescent obesity prevention literature, particularly in non-Western contexts. The Green Apple program's success suggests that weight-neutral, culturally adapted interventions can effectively promote health and well-being among Saudi female adolescents. Its impact on increasing MNCD-related knowledge and its promising trends in reducing sedentary behavior, body image discrepancy, and eating disordered symptoms reinforce the value of moving away from weight-centric approaches to obesity prevention.

Overall, these findings add to the growing evidence in support of weight-inclusive health promotion strategies and underscore the need for culturally appropriate interventions in addressing global health challenges.

Conclusions

This study demonstrated that the Green Apple program, a school-based intervention grounded in a non–weight-centric approach, effectively enhanced adolescents' health behaviors and psychosocial outcomes. Specifically, the enhanced intervention (Green Apple+MNCD) led to significant improvements in knowledge about MNCDs across all schools. In addition, the Green Apple program without the MNCD session resulted in a notable reduction in sedentary behavior by −3.12 hours per day among rural students and a significant overall reduction of −2.39 hours per day across rural and urban schools. Both versions of the program showed promising trends in improving body image discrepancy and disordered eating symptoms without causing harm. At follow-up, a statistically significant difference was observed in favor of the enhanced intervention (Green Apple + MNCD) group, indicating a greater reduction in body image discrepancy compared with the Green Apple group. Furthermore, in urban schools, only the enhanced intervention (Green Apple+MNCD) group demonstrated a significant reduction in disordered eating symptoms. These findings support the effectiveness and safety of implementing culturally tailored, weight-neutral interventions in adolescent populations.

Future research should pursue longitudinal studies to assess the sustainability of knowledge gains and behavioral changes over extended periods. RCTs are recommended to confirm the program's effectiveness in broader and more diverse populations, including male adolescents and students in other regions of Saudi Arabia. In addition, implementation research is needed to evaluate the program's scalability, feasibility, and cost-effectiveness in real-world school settings. Specific priority areas include adapting the intervention to different cultural or socioeconomic contexts, integrating digital tools to enhance reach and engagement, and examining the role of school infrastructure and policy in supporting long-term behavior change. For instance, a mobile app could track daily activity and provide visual feedback on sedentary behavior, while an online platform could host culturally tailored modules on nutrition and chronic disease prevention, accessible to both students and parents. These digital enhancements would not only broaden the program's reach but also increase engagement and retention of health messages across diverse student populations. Addressing these areas will help inform national strategies aimed at preventing metabolic chronic diseases through inclusive, sustainable, and culturally sensitive school-based interventions.

Acknowledgments

The authors would like to express sincere gratitude to the Ministry of Education School Health Department in Medina and all the participating schools for their invaluable support.

Grammarly (Grammarly Inc), an artificial intelligence–powered writing support tool, was used to check grammar, spelling, and language clarity. No content, data analysis, or scientific interpretation was generated by artificial intelligence. All study content and conclusions were developed and verified by the authors.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: AZ, FA, MA
Data curation: AZ
Formal analysis: AZ
Investigation: AZ, RJ, NA, EM, BA, MA
Methodology: AZ
Project administration: AZ, RJ
Resources: MA
Software: AZ
Supervision: AZ
Validation: AZ, MA
Visualization: AZ
Writing—original draft preparation: AZ
Writing—review and editing: AZ, FA, RJ, NA, MA
All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details on data completeness.
[DOCX File, 16 KB - i-jmr_v14i1e71341_app1.docx]

Multimedia Appendix 2
Baseline characteristics between types of intervention.
[DOCX File, 16 KB - i-jmr_v14i1e71341_app2.docx]

Multimedia Appendix 3
Details on types of units included within each intervention.
[DOCX File, 15 KB - i-jmr_v14i1e71341_app3.docx]

Checklist 1
TREND Statement checklist.
[DOCX File, 23 KB - i-jmr_v14i1e71341_app4.docx]

References

1. Rubino F, Puhl RM, Cummings DE, et al. Joint international consensus statement for ending stigma of obesity. Nat Med 2020 Apr;26(4):485-497. [doi: 10.1038/s41591-020-0803-x] [Medline: 32127716]
2. Puhl RM, Lessard LM, Larson N, Eisenberg ME, Neumark-Stzainer D. Weight stigma as a predictor of distress and maladaptive eating behaviors during COVID-19: longitudinal findings from the EAT study. Ann Behav Med 2020 Oct 1;54(10):738-746. [doi: 10.1093/abm/kaaa077] [Medline: 32909031]
3. Lin YC, Lin CY, Saffari M, et al. Weight stigma is associated with body mass index among college students in Taiwan: the mediated role of internalized weight stigma. BMC Psychol 2023 Nov 1;11(1):365. [doi: 10.1186/s40359-023-01414-w] [Medline: 37915074]
4. Nutter S, Ireland A, Alberga AS, et al. Weight bias in educational settings: a systematic review. Curr Obes Rep 2019 Jun;8(2):185-200. [doi: 10.1007/s13679-019-00330-8] [Medline: 30820842]
5. Levinson JA, Kinkel-Ram S, Myers B, Hunger JM. A systematic review of weight stigma and disordered eating cognitions and behaviors. Body Image 2024 Mar;48:101678. [doi: 10.1016/j.bodyim.2023.101678] [Medline: 38278088]
6. Dugmore JA, Winten CG, Niven HE, Bauer J. Effects of weight-neutral approaches compared with traditional weight-loss approaches on behavioral, physical, and psychological health outcomes: a systematic review and meta-analysis. Nutr Rev 2020 Jan 1;78(1):39-55. [doi: 10.1093/nutrit/nuz020]
7. Major B, Rathbone JA, Blodorn A, Hunger JM. The countervailing effects of weight stigma on weight-loss motivation and perceived capacity for weight control. Pers Soc Psychol Bull 2020 Sep;46(9):1331-1343. [doi: 10.1177/0146167220903184] [Medline: 32046597]

XSL•FO
RenderX

8. Hunger JM, Dodd DR, Smith AR. Weight discrimination, anticipated weight stigma, and disordered eating. Eat Behav 2020 Apr;37:101383. [doi: 10.1016/j.eatbeh.2020.101383] [Medline: 32438268]
9. Motevalli M, Drenowatz C, Tanous DR, Khan NA, Wirnitzer K. Management of childhood obesity-time to shift from generalized to personalized intervention strategies. Nutrients 2021 Apr 6;13(4):1200. [doi: 10.3390/nu13041200] [Medline: 33917383]
10. Nutter S, Eggerichs LA, Nagpal TS, et al. Changing the global obesity narrative to recognize and reduce weight stigma: a position statement from the World Obesity Federation. Obes Rev 2024 Jan;25(1):e13642. [doi: 10.1111/obr.13642] [Medline: 37846179]
11. AlEnazi S, AlAjlan R, AlKhalaf H, et al. Prevalence of obesity among children and adolescents in Saudi Arabia: a multicenter population-based study. Saudi J Med Med Sci 2023;11(1):19-25. [doi: 10.4103/sjmms.sjmms_417_22] [Medline: 36909009]
12. Alhamed A, Al-Zeer A, Alsaawi F, et al. The shifting patterns of childhood obesity: insights from national school screening data. Saudi Pharm J 2023 Dec;31(12):101862. [doi: 10.1016/j.jsps.2023.101862] [Medline: 38028221]
13. Al-Hussaini A, Bashir MS, Khormi M, et al. Overweight and obesity among Saudi children and adolescents: where do we stand today? Saudi J Gastroenterol 2019;25(4):229-235. [doi: 10.4103/sjg.SJG_617_18] [Medline: 31187784]
14. Almubark RA, Alqahtani SA. Epidemiology of obesity and control interventions in Saudi Arabia. East Mediterr Health J 2023 Dec 21;29(12):987-994. [doi: 10.26719/emhj.23.086] [Medline: 38279867]
15. Al Eid AJ, Alahmed ZA, Al-Omary SA, Alharbi SM. RASHAKA Program: a collaborative initiative between Ministry of Health and Ministry of Education to control childhood obesity in Saudi Arabia. Saudi J Obesity 2017;5(1):22. [doi: 10.4103/sjo.sjo_5_17]
16. Banany M, Kang M, Gebel K, Sibbritt D. A systematic review of school-based weight-related interventions in the Gulf Cooperation Council countries. Syst Rev 2024 Feb 14;13(1):66. [doi: 10.1186/s13643-024-02475-7] [Medline: 38355590]
17. Al-Daghri NM, Wani K, Khattak MNK, et al. Cardiometabolic effects of a 12-month, COVID-19 lockdown-interrupted lifestyle education program for Arab adolescents. Front Pediatr 2022;10:887138. [doi: 10.3389/fped.2022.887138] [Medline: 35783315]
18. Abdulwahab IJ, Alzahrani NA, Khouja JH. A comprehensive review of school-based lifestyle interventions to reduce childhood obesity in the Eastern Mediterranean Region. Cureus 2024 Jul;16(7):e64305. [doi: 10.7759/cureus.64305] [Medline: 39130893]
19. Ahmad Bahathig A, Abu Saad H. The effects of a physical activity, nutrition, and body image intervention on girls in intermediate schools in Saudi Arabia. Int J Environ Res Public Health 2022 Sep 8;19(18):11314. [doi: 10.3390/ijerph191811314] [Medline: 36141585]
20. Kenney EL, Wintner S, Lee RM, Austin SB. Obesity prevention interventions in US public schools: are schools using programs that promote weight stigma? Prev Chronic Dis 2017 Dec 28;14:E142. [doi: 10.5888/pcd14.160605] [Medline: 29283880]
21. Ali AA, Aqeel AA, Shami MO, et al. Relationship between depression, anxiety, stress, and weight self-stigma among youths in Saudi Arabia: a nationwide study. Cureus 2024 Feb;16(2):e54125. [doi: 10.7759/cureus.54125] [Medline: 38487156]
22. Alrougui K, Alshammari GM, Alrwis KN, Alabdulkarem KB, Yahya MA. Nutritional status and associated socioeconomic factors among low-income Saudi females. Front Sustain Food Syst 2025;9. [doi: 10.3389/fsufs.2025.1510426]
23. Addas A. Impact of neighborhood safety on adolescent physical activity in Saudi Arabia: gender and socio-economic perspectives. Front Public Health 2025;13:1520851. [doi: 10.3389/fpubh.2025.1520851] [Medline: 39916720]
24. Almaqhawi A. Perceived barriers and facilitators of physical activity among Saudi Arabian females living in the East Midlands. J Taibah Univ Med Sci 2022 Jun;17(3):384-391. [doi: 10.1016/j.jtumed.2021.11.002] [Medline: 35722239]
25. Des Jarlais DC, Lyles C, Crepaz N, TREND Group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Public Health 2004 Mar;94(3):361-366. [doi: 10.2105/ajph.94.3.361] [Medline: 14998794]
26. Chang TH, Stuart EA. Propensity score methods for observational studies with clustered data: a review. Stat Med 2022 Aug 15;41(18):3612-3626. [doi: 10.1002/sim.9437] [Medline: 35603766]
27. Langworthy B, Wu Y, Wang M. An overview of propensity score matching methods for clustered data. Stat Methods Med Res 2023 Apr;32(4):641-655. [doi: 10.1177/09622802221133556] [Medline: 36426585]
28. Champion VL, Skinner CS. The Health Belief Model, 4th edition: Jossey-Bass; 2008.
29. Bussey K, Bandura A. Social cognitive theory of gender development and differentiation. Psychol Rev 1999 Oct;106(4):676-713. [doi: 10.1037/0033-295x.106.4.676] [Medline: 10560326]
30. Chang MW, Nitzke S, Brown R, Resnicow K. A community based prevention of weight gain intervention (Mothers In Motion) among young low-income overweight and obese mothers: design and rationale. BMC Public Health 2014 Mar 25;14(1):280. [doi: 10.1186/1471-2458-14-280] [Medline: 24666633]
31. Post RC. A new approach to dietary guidelines communications: make MyPlate, your plate. Child Obes 2011 Oct;7(5):349-351. [doi: 10.1089/chi.2011.0500.post]
32. Wansink B, Kranz S. Who's using MyPlate? J Nutr Educ Behav 2013;45(6):728-732. [doi: 10.1016/j.jneb.2013.03.009] [Medline: 23791898]

XSL•FO
RenderX

33. Zuair A, Alhowaymel FM, Jalloun RA, et al. Body fat and obesity rates, cardiovascular fitness, and the feasibility of a low-intensity non-weight-centric educational intervention among late adolescents: quasi-experimental study. JMIR Pediatr Parent 2025 Jan 24;8:e67213. [doi: 10.2196/67213] [Medline: 39865566]
34. Al-Hazzaa HM, Musaiger AO, Group AR. Arab Teens Lifestyle Study (ATLS): objectives, design, methodology and implications. DMSO 2011;4:417. [doi: 10.2147/DMSO.S26676]
35. Al-Hazzaa HM, Al-Sobayel HI, Musaiger AO. Convergent validity of the Arab Teens Lifestyle Study (ATLS) physical activity questionnaire. Int J Environ Res Public Health 2011 Sep;8(9):3810-3820. [doi: 10.3390/ijerph8093810] [Medline: 22016718]
36. Alahmadi MA, Almasoud KH, Aljahani AH, et al. Validity and reliability of the Arabic sedentary behavior questionnaire among university students aged between 18-30 years old. BMC Public Health 2023 Jan 18;23(1):128. [doi: 10.1186/s12889-023-15030-1] [Medline: 36653773]
37. Chau JY, Grunseit AC, Chey T, et al. Daily sitting time and all-cause mortality: a meta-analysis. PLoS ONE 2013;8(11):e80000. [doi: 10.1371/journal.pone.0080000] [Medline: 24236168]
38. Alahmadi MA, Almasoud KH, Aljahani AH, et al. The prevalence of sedentary behavior among university students in Saudi Arabia. BMC Public Health 2024 Feb 26;24(1):605. [doi: 10.1186/s12889-024-18107-7] [Medline: 38408947]
39. Marías Y, Glasauer P. Guidelines for Assessing Nutrition-Related Knowledge, Attitudes and Practices: Food and Agriculture Organization of the United Nations (FAO); 2014.
40. Simeon DT, Rattan RD, Panchoo K, Kungeesingh KV, Ali AC, Abdool PS. Body image of adolescents in a multi-ethnic Caribbean population. Eur J Clin Nutr 2003 Jan;57(1):157-162. [doi: 10.1038/sj.ejcn.1601515] [Medline: 12548311]
41. Kronenfeld LW, Reba-Harrelson L, Von Holle A, Reyes ML, Bulik CM. Ethnic and racial differences in body size perception and satisfaction. Body Image 2010 Mar;7(2):131-136. [doi: 10.1016/j.bodyim.2009.11.002] [Medline: 20096656]
42. Kelly NR, Bulik CM, Mazzeo SE. An exploration of body dissatisfaction and perceptions of Black and White girls enrolled in an intervention for overweight children. Body Image 2011 Sep;8(4):379-384. [doi: 10.1016/j.bodyim.2011.05.003] [Medline: 21700518]
43. Latiff AA, Muhamad J, Rahman RA. Body image dissatisfaction and its determinants among young primary-school adolescents. J Taibah Univ Med Sci 2018 Feb;13(1):34-41. [doi: 10.1016/j.jtumed.2017.07.003] [Medline: 31435300]
44. Albawardi NM, AlTamimi AA, AlMarzooqi MA, Alrasheed L, Al-Hazzaa HM. Associations of body dissatisfaction with lifestyle behaviors and socio-demographic factors among Saudi females attending fitness centers. Front Psychol 2021;12:611472. [doi: 10.3389/fpsyg.2021.611472] [Medline: 33643138]
45. Aoun A, Azzam J, Jabbour FE, et al. Validation of the Arabic version of the SCOFF questionnaire for the screening of eating disorders. East Mediterr Health J 2015 Aug 27;21(5):326-331. [doi: 10.26719/2015.21.5.326] [Medline: 26343121]
46. Al Mestaka N, Alneyadi A, AlAhbabi A, AlMatrushi A, AlSaadi R, Alketbi LB. Prevalence of probable eating disorders and associated risk factors in children and adolescents aged 5-16 years in Al Ain City, United Arab Emirates: observational case-control study. J Eat Disord 2023 Jul 7;11(1):114. [doi: 10.1186/s40337-023-00840-w] [Medline: 37420301]
47. Schafer JL, Yucel RM. Computational strategies for multivariate linear mixed-effects models with missing values. J Comput Graph Stat 2002 Jun;11(2):437-457. [doi: 10.1198/106186002760180608]
48. Hsu YT, Buckworth J, Focht BC, O'Connell AA. Feasibility of a self-determination theory-based exercise intervention promoting healthy at every size with sedentary overweight women: Project CHANGE. Psychol Sport Exerc 2013 Mar;14(2):283-292. [doi: 10.1016/j.psychsport.2012.11.007]
49. Bagherniya M, Taghipour A, Sharma M, et al. Obesity intervention programs among adolescents using social cognitive theory: a systematic literature review. Health Educ Res 2018 Feb 1;33(1):26-39. [doi: 10.1093/her/cyx079]
50. Chew HSJ, Chng S, Rajasegaran NN, Choy KH, Chong YY. Effectiveness of acceptance and commitment therapy on weight, eating behaviours and psychological outcomes: a systematic review and meta-analysis. Eat Weight Disord 2023 Feb 10;28(1):6. [doi: 10.1007/s40519-023-01535-6] [Medline: 36763199]
51. Jacob CM, Hardy-Johnson PL, Inskip HM, et al. A systematic review and meta-analysis of school-based interventions with health education to reduce body mass index in adolescents aged 10 to 19 years. Int J Behav Nutr Phys Act 2021 Jan 4;18(1):1. [doi: 10.1186/s12966-020-01065-9] [Medline: 33397403]
52. Puhl RM, Lessard LM, Pearl RL, Himmelstein MS, Foster GD. International comparisons of weight stigma: addressing a void in the field. Int J Obes (Lond) 2021 Sep;45(9):1976-1985. [doi: 10.1038/s41366-021-00860-z] [Medline: 34059785]
53. Heitmann BL, Køster-Rasmussen R, Meyer LB, et al. Debating weight loss vs. weight neutral strategies for improvements of health. Curr Obes Rep 2024 Dec;13(4):832-842. [doi: 10.1007/s13679-024-00587-8] [Medline: 39289256]
54. Hoare JK, Lister NB, Garnett SP, Baur LA, Jebeile H. Weight-neutral interventions in young people with high body mass index: a systematic review. Nutr Diet 2023 Feb;80(1):8-20. [doi: 10.1111/1747-0080.12729] [Medline: 35411702]
55. Weigensberg MJ, Lane CJ, Ávila Q, et al. Imagine HEALTH: results from a randomized pilot lifestyle intervention for obese Latino adolescents using Interactive Guided ImagerySM. BMC Complement Altern Med 2014 Jan 17;14(1):28. [doi: 10.1186/1472-6882-14-28] [Medline: 24433565]
56. Suárez R, Cucalon G, Herrera C, et al. Effects of health at every size based interventions on health-related outcomes and body mass, in a short and a long term. Front Nutr 2024;11:1482854. [doi: 10.3389/fnut.2024.1482854]

XSL•FO
RenderX

57. Penney TL, Kirk SFL. The health at every size paradigm and obesity: missing empirical evidence may help push the reframing obesity debate forward. Am J Public Health 2015 May;105(5):e38-e42. [doi: 10.2105/AJPH.2015.302552] [Medline: 25790393]
58. Bookari K. What is the level of nutrition literacy of Saudi adolescents? A national wide exploratory cross-sectional study. Front Nutr 2022;9:1113910. [doi: 10.3389/fnut.2022.1113910]
59. Al-Hazzaa HM, Albawardi NM. Obesity, lifestyle behaviors, and dietary habits of Saudi adolescents living in Riyadh (ATLS-2 Project): revisited after a ten-year period. Life (Basel) 2021 Oct 13;11(10):1078. [doi: 10.3390/life11101078] [Medline: 34685449]
60. Puhl RM, Himmelstein MS, Pearl RL. Weight stigma as a psychosocial contributor to obesity. Am Psychol 2020;75(2):274-289. [doi: 10.1037/amp0000538] [Medline: 32053000]
61. Błaszczyk-Bębenek E, Jagielski P, Nowak R, Ostachowska-Gąsior A. Dissatisfaction and discrepancy in body-image perceptions among adolescents aged 10-17 years from rural areas of southern Poland - a cross-sectional study. Ann Agric Environ Med 2025 Jun 27;32(2):240-247. [doi: 10.26444/aaem/193461] [Medline: 40586501]
62. Zuair A. Effect of the reassured self-compassion-based school program on anxiety, video game addiction, and body image among rural female adolescents: retrospective study. JMIR Form Res 2025 Feb 19;9:e68840. [doi: 10.2196/68840] [Medline: 39970430]
63. Damasceno VDO, De Lima DRAA, De Brito MJC, et al. Psychometric properties of a silhouette scale on body image in adults. J Phys Educ Sport 2020;20(3):1295-1301. [doi: 10.7752/jpes.2020.03180]
64. Emeir M, A. Alshammai. Latifa M. Al Hamdan M, S. Almutairi B, Eid Alshammari M, Alshammari N. Transforming lifestyles for healthier communities: a literature review and strategic vision for the Hail Region, Saudi Arabia. JPM 2024 Jan 29;12(1):7-12. [doi: 10.12691/jpm-12-1-2]
65. Madsen KA, Thompson HR, Linchey J, et al. Effect of school-based body mass index reporting in California public schools: a randomized clinical trial. JAMA Pediatr 2021 Mar 1;175(3):251-259. [doi: 10.1001/jamapediatrics.2020.4768] [Medline: 33196797]
66. Tingle E, Saunders JF, Nutter S, Russell-Mayhew S. Taking weight out of the equation: unintended harms of weight-focused health discourse in schools. Journal of Physical Education, Recreation & Dance 2023 Feb 12;94(2):49-58. [doi: 10.1080/07303084.2022.2146818] [Medline: 18858293]
67. Richmond TK, Thurston IB, Sonneville KR. Weight-focused public health interventions-no benefit, some harm. JAMA Pediatr 2021 Mar 1;175(3):238-239. [doi: 10.1001/jamapediatrics.2020.4777] [Medline: 33196761]
68. Lessard LM, Puhl RM. Reducing educators' weight bias: the role of school-based anti-bullying policies. J Sch Health 2021 Oct;91(10):796-801. [doi: 10.1111/josh.13068] [Medline: 34426980]
69. Kataria I, Jackson-Morris A, Jewell J, et al. Weight stigma among adolescents in three low- and middle-income countries. J Glob Health 2022 Dec 16;12:04098. [doi: 10.7189/jogh.12.04098] [Medline: 36520445]
70. Watson D, Hughes K, Robinson E, Billette J, Bombak AE. Patient recommendations for providers to avoid stigmatizing weight in rural-based women with low income. J Patient Cent Res Rev 2021;8(1):20-30. [doi: 10.17294/2330-0698.1752] [Medline: 33511250]
71. Zuair AA. Assessment of a school-based, nursing-lead program to combat iron deficiency anemia among Saudi female adolescents: a pilot exploratory study. Discov Public Health 2025;22(1):47. [doi: 10.1186/s12982-025-00427-7]

Abbreviations

ATLS: Arab Teens Lifestyle Study
HAES: Health at Every Size
HBM: health belief model
MET: metabolic equivalent
MNCD: metabolic noncommunicable disease
RCT: randomized controlled trial
SCT: social cognitive theory
TTM: transtheoretical model

XSL•FO
RenderX

Edited by N Cahill; submitted 15.01.25; peer-reviewed by R Nooripour, T Baranowski; revised version received 06.09.25; accepted 09.09.25; published 22.10.25.

Please cite as:
Zuair A, Jalloun R, Alzahrani N, Alhowaymel F, Merza E, Alhumaidi B, Alahmadi M
Evaluating the Safety and Efficacy of a Non–Weight-Centric Approach to Obesity Prevention in Rural and Urban Female Adolescents: Quasi-Experimental Study
Interact J Med Res 2025;14:e71341
URL: https://www.i-jmr.org/2025/1/e71341
doi:10.2196/71341

© Areeg Zuair, Rola Jalloun, Naif Alzahrani, Fahad Alhowaymel, Esraa Merza, Bandar Alhumaidi, Mohammad Alahmadi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 22.10.2025. 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

Simplified Medication Adherence Questionnaire (SMAQ) for People Living With HIV in a National Hospital in Mexico: Instrument Validation Study

Luis Eduardo Del Moral Trinidad[1], MPH; Luz Alicia González Hernández[2], PhD; Jaime Federico Andrade Villanueva[2], MCR; Pedro Martínez-Ayala[2], MCR; Adriana Valle Rodríguez[2], MD; Vida Veronica Ruíz Herrera[2], MD; José Adán Vizcaíno Résendiz[2], MOP; Melva Guadalupe Herrera Godina[3], PhD; Sergio Dominguez-Lara[4], PhD

[1]Department of Public Health Sciences, University of Guadalajara, Guadalajara, Mexico

[2]HIV Unit, Hospital Civil de Guadalajara, Guadalajara, Mexico

[3]Institute of Biomedical Sciences Research, University of Guadalajara, Guadalajara, Mexico

[4]Research Institute, Facultad de Ciencias de la Comunicación, Turismo y Psicología, University of San Martín de Porres, Lima, Peru

Corresponding Author:
Sergio Dominguez-Lara, PhD
Research Institute
Facultad de Ciencias de la Comunicación, Turismo y Psicología
University of San Martín de Porres
Avenue Tomás Marsano 242 (5th floor)
Lima, 15048
Peru
Phone: 51 5136300
Email: sdominguezmpcs@gmail.com

Abstract

Background: Adherence to antiretroviral therapy is a critical component in achieving viral suppression in people living with HIV in addition to increasing overall quality of life. Several indirect methods have been used to measure adherence including the Simplified Medication Adherence Questionnaire (SMAQ).

Objective: The objective of this study is to evaluate the reliability and validity of the SMAQ in men living with HIV/AIDS attending a Mexican national hospital.

Methods: A cross-sectional analytical design study was carried out in a Mexican National Hospital in Jalisco, including men aged >18 years with at least 3 months of antiretroviral treatment, excluding those with cognitive difficulties in answering the survey. A minimum sample size was calculated to detect the contribution of the variables within the model. The analysis included descriptive tests, confirmatory factor analysis, reliability and validity assessment, correlation between adherence and viral load, and association between viral load and adherence.

Results: The final analysis included a total of 260 patients with a mean age of 43 (SD 12) years and an average of 8.97 (SD 6.33) years on antiretroviral treatment. The SMAQ showed sufficient structural validity (comparative fit index=1, root-mean-square error of approximation=0, 90% CI 0-0.085) with satisfactory factor loadings on most questions except item 2 (Do you always take your medication at the prescribed time?). The reliability of the scale is acceptable (Cronbach α=0.702, ω=0.718). Adherence correlated with viral load significantly but not with recent TCD4 lymphocyte levels. Patients classified as adherent were three times more likely to be undetectable than nonadherent patients (odds ratio 3.31, 95% CI 1.13-9.64, *P*=.04).

Conclusions: The SMAQ represents an adequate tool to assess adherence in men living with HIV in the Mexican context, this will contribute to this study and compression of adherence to establish future intervention programs.

(Interact J Med Res 2025;14:e59562) doi:10.2196/59562

KEYWORDS

treatment adherence; HIV; Mexico; validation; Spanish; Hispanic; cross sectional; surveys; questionnaires; scales; adherence; viral load; sexually transmitted infection; STI; drugs; pharmacotherapy; medication; simplified medication adherence questionnaire; SMAQ

XSL•FO
RenderX

Introduction

According to the Joint United Nations Program on HIV/AIDS, 39 million people were living with HIV by 2022, of which only 76.4% were receiving antiretroviral treatment (ART). In Mexico, by 2022, there were 270,000 cases registered with the Ministry of Health, 80% (n=270,000) of which were men; in addition, Jalisco ranks 4th in prevalence of people living with HIV with a record of 7134 patients on ART and 78% of this population has achieved viral suppression through such treatment [1,2].

The World Health Organization defines adherence to treatment as an individual's behavior regarding medications, diet, and lifestyle changes that correspond to the recommendations provided by a health professional [3].

Therapeutic adherence is a complex process that is made up of a personal component represented by the patient, where their attitudes toward their disease and the positions they take on it are concentrated, as well as a relational component involving the health professional and the health structure that surrounds them. All these components work synergistically toward a common goal to benefit the patient's health [4].

Adherence to ART is a critical component to achieve viral suppression in people living with HIV as well as to increase the overall quality of life [5]. Its study is important because in Mexico there has been a sustained prevalence in recent years, and it is necessary to improve the tools available to optimize treatment success [6].

There are various methods for measuring adherence in a patient; however, there is no gold standard for this purpose. Due to this, its measurement will depend on the characteristics of the population studied; in addition, the method should have basic psychometric standards of acceptable validity and reliability [7]. These methods are classified as direct, those that directly quantify the drug and its metabolites in blood or any other fluid or tissue, but they are costly and impractical for routine implementation, and indirect, that is they base their measurement on pill counts and self-reports, among others [8].

An indirect tool that has been used by several authors is the Simplified Medication Adherence Questionnaire (SMAQ). This questionnaire has been used in studies to evaluate adherence to ART in people living with HIV, and its six-question structure makes it practical in clinical contexts where a rapid evaluation is required [9,10]. This scale was introduced in 2002 by Knobel et al [9]. They designed the scale intending to create a questionnaire to identify nonadherent patients and found that this instrument had a sensitivity of 72%, a specificity of 91%, and a likelihood ratio of 7.94, as well as a Cronbach α of 0.75 [9].

The SMAQ questionnaire has also been validated in different pathologies, medical conditions, and chronic diseases, such as hypertension, diabetes, and tuberculosis. These validations have shown consistent results, supporting the usefulness and validity of the questionnaire in different health care settings [9,11]. The instrument was originally developed for the Spanish population in 1999, but the current social and pharmacological context differs for the population we are now studying. It is important to evaluate the validity and reliability of the instrument when used in contexts different from its original development, to ensure the quality of the collected information. This is because the metric quality of a self-report questionnaire must be explored in the context where it will be applied. Otherwise, the psychometric properties would be compromised, leading to negative consequences for the evaluation [12,13].

This questionnaire is not yet validated in Latin people, thus, its utility in our population is unknown. There is only a pilot study with 10 participants that used the SMAQ in Peruvian people living with HIV; where item comprehension was evaluated, as cultural applicability and social acceptance [14]. To our knowledge, no studies have been conducted to validate this instrument in the Mexican population. This study aimed to evaluate the reliability and validity of the SMAQ in people living with HIV/AIDS attending a Mexican national hospital.

Methods

Study Design

This study is a cross-sectional analytical design study concerning people living with HIV/AIDS receiving ART at the Civil Hospital of Guadalajara "Fray Antonio Alcalde."

Selection of Participants

Men older than 18 years of age receiving ART for at least three months before their inclusion in this study were included. Every participant gave his informed consent before participating in the research. On the other hand, those persons diagnosed with serious mental illnesses that may affect their ability to understand or answer the questionnaire were excluded, as well as those with cognitive or communication difficulties that may hinder their participation in the evaluation. Additionally considered as an exclusion criterion is not having performed the test to detect viral load within the period established for this study.

Sample Size Calculation

To determine the sample size, we used a calculator that used a structural equation model approach, in which we anticipated an effect size (factor loading) of 0.5, which would reflect a significant contribution of the latent construct to the observed variables (items), ensuring sufficient construct reliability, a power of 80%, one latent variable, five observable variables, and a probability level of 0.05, resulting in a minimum sample of 100 participants, to detect the specified effect given the structural complexity of this model [15,16].

Variables

The variables considered for this study were age in years, schooling (considered as the last completed grade of studies); municipality of residence (with the category "other" for those participants who indicated not being from a nearby municipality); marital status; employment (was considered positive when they would indicate having an economically remunerated activity); the use of tobacco, alcohol, and illicit drugs (it was considered positive when patients indicated to use in the last 30 days on more than two occasions); the number of pills (including those belonging to a treatment other than ART);

XSL•FO
RenderX

time living with HIV and time on ART (which were calculated considering from the date of HIV diagnosis and the start of ART until the date of enrollment in this study); and finally, clinical stage was classified according to the Center of Disease Control classification [17].

Instrument

The questionnaire consists of 6 questions: the first 4 can be answered with "yes" or "no," while the last 2 require numerical responses: "1. Do you ever forget to take your medication?" "2. Do you always take your medication at the prescribed time?" "3. Do you ever stop taking the drugs if you feel unwell?" "4. Did you forget to take the medication over the weekend?" "5. In the last week, how many times did you miss a dose?" and "6. In the last 3 months, how many full days did you miss taking the medication?" The six questions assess three components of ART adherence: (1) the intentional (question 3), (2) the unintentional (questions 1 and 2), and (3) frequency or quantity (questions 4, 5, and 6). The patient is classified as nonadherent if they answer any in a "nonadherence sense" and if they report missing more than two doses in the last week or reports not having taken more than 2 full days of medication in the last three months [10].

Statistical Analysis

For descriptive analysis, normality tests were performed as necessary. Quantitative variables were described with means or medians, and qualitative variables with frequencies and percentages.

The evidence of validity based on internal structure. was assessed by confirmatory factor analysis using the weighted least squares method with adjusted mean and variance [18]. Model fit was assessed using recognized indices, such as comparative fit index (CFI>0.90) [19], root-mean-square error of approximation (RMSEA<0.08) [20], and weight root-mean-square residual (WRMR<1) [21].

To determine reliability, we estimated the internal consistency coefficients according to Ponterotto and Charter [22]. Scores Cronbach α>0.70 were considered as reliable scores [22]. In addition, we investigated validity by examining the relationship between adherence and undetectable plasmatic viral load using the Pearson correlation coefficient [23].

Finally, Fisher exact tests were used to calculate the odds ratios (OR) to evaluate the association between having an undetectable (patient in control) versus detectable (patient not in control) plasmatic viral load and the scale classification of adherent versus nonadherent (based on the concept that a patient with an undetectable viral load was an adherent patient) [5].

The descriptive and correlational analyses were carried out in SPSS (version 24; IBM Corp) software, while the factor analysis was performed with Mplus (version 7; Muthén & Muthén) software [19]. Finally, the analysis of ORs was performed in Epi Info (Centers for Disease Control and Prevention) software in the Stat Cal module [24].

Ethical Considerations

This research was approved by the Bioethics Committee of the Civil Hospital of Guadalajara (169/23), and it was carried out under the ethical standards for research on human participants. All participants signed an informed consent form that outlined the objectives of the research, the procedures to be followed, and the contact information for the research team. Any questions or concerns were addressed verbally prior to signing the consent form. The information collected was handled exclusively by the research team, and to protect participants' identities an alphanumeric code was assigned to each. Additionally, identifying data were excluded from the database used for analysis. Finally, no direct financial compensation was provided to participants for their involvement in the study.

Results

Characteristics of the Participants

A total of 299 patients were evaluated, 259 men and 40 women, and due to the ratio between groups, we decided to exclude the group of women from the analysis, to maintain the homogeneity and representativeness of the sample [25].

Of the 259 participants included, the mean age was 43 (SD 12) years with an average of 8.97 (SD 6.33) years in ART treatment. Considering schooling, the majority had high school (n=87, 34.6%) or middle school (n=57, 22%) education. The most common municipality of residence was Guadalajara (n=110, 42.5%) and the vast majority were single (n=214, 82.6%). Regarding substance use, 89.6% (n=232) had not consumed illicit substances in the last 30 days while 33.6% (n=87) had consumed alcohol and tobacco in the same period. The mean number of pills per day was 2 (SD 2) and the mean number of years living with HIV was 10 (SD 7.17) with an average CD4+ T lymphocyte count of 686.5 cells/µL (SD 353.88; Table 1).

XSL•FO
RenderX

Table 1. Sociodemographic characteristics of the sample (N=259).

	Values
Age (years), mean (SD)	43 (12)
Schooling, n (%)	
Illiterate	2 (0.77)
Can read and write	16 (6.2)
Primary school	44 (17)
Junior high school	57 (22)
High school	87 (34.6)
University	50 (19.3)
Postgraduate	3 (1.2)
Municipality of residence, n (%)	
Guadalajara	110 (42.5)
Zapopan	32 (12.4)
Tlaquepaque	24 (9.3)
Tonalá	16 (6.2)
Zapotlanejo	2 (0.8)
Tepatitlán	1 (0.4)
La Barca	1 (0.4)
Arandas	2 (0.8)
Zapotlán el Grande	2 (0.8)
Other	69 (26.7)
Marital status, n (%)	
Single	214 (82.6)
Married	16 (6.2)
Common law marriage	22 (8.5)
Widower	2 (0.8)
Regular employment, n (%)	
No	37 (14.3)
Yes	222 (85.7)
Used any illicit substance within the last 30 days, n (%)	
No	232 (89.6)
Yes	27 (10.4)
Consumed alcohol in the last 30 days, n (%)	
No	172 (66.4)
Yes	87 (31.3)
Used tobacco in the last 30 days	
No, n (%)	171 (66)
Yes, n (%)	87 (33.6)
Pills taken per day, mean (SD)	2 (2)
Years in ART[a], mean (SD)	8.97 (6.33)
Years living with HIV, mean (SD)	10.08 (7.17)
CD4+ T lymphocyte count, mean (SD)	686.47 (353.88)

XSL•FO
RenderX

	Values
Clinical stage according to the CDC[b], n (%)	
A1	122 (47.1)
A2	20 (7.7)
A3	4 (1.4)
B1	16 (6.2)
B2	4 (1.5)
B3	4 (1.5)
C1	37 (14.3)
C2	32 (12.4)
C3	20 (7.7)

[a]ART: antiretroviral therapy.

[b]CDC: Centers for Disease Control and Prevention.

Responses From the Participants

Regarding the responses to the questionnaire, 66.6% (n=172) of the respondents reported not forgetting their medication, while 33.4% (n=87) admitted to having forgotten it at least once. On the second question, 80.3% (n=198) stated that they took their medication at the indicated time and 19.7% (n=61) admitted not following this pattern. Regarding whether they stopped taking the drugs if they felt unwell, the great majority (n=242, 92.2%) stated that they did not do so but 7.8% (n=17) admitted having done so at some time.

The majority reported not having forgotten to take their medication over the weekend (n=250, 96.6%). In the last week, the majority (n=242, 93.8%) reported missing only one dose, while 6.2% admitted missing a dose once or twice.

On the other hand, of the responses to question 6, a total of 66.5% (n=174) indicated not having missed taking medication on any full day, while 30.6% (n=174) acknowledged having experienced some degree of medication interruption.

Finally, according to the scores of the questionnaire, 66.6% (n=174) of participants reported being adherent while 33.4% (n=88) reported being nonadherent (Table 2).

XSL•FO
RenderX

Table 2. Distribution of responses to the Simplified Medication Adherence Questionnaire (N=259).

Questions	Participants, n (%)
1. Do you ever forget to take your medication?	
No	172 (66.6)
Yes	87 (33.4)
2. Do you always take your medication at the prescribed time?	
No	61 (19.7)
Yes	198 (80.3)
3. Do you ever stop taking the drugs if you feel unwell?	
No	242 (92.2)
Yes	17 (7.8)
4. Did you forget to take your medication over the weekend?	
No	250 (96.6)
Yes	9 (3.4)
5. In the last week, how many times did you not take a dose?	
1	242 (93.8)
2	16 (5.9)
3	1 (0.3)
6. In the last 3 months, how many full days did you not take your medication?	
0	174 (66.5)
1	51 (20.6)
2	17 (5.9)
3	7 (3.1)
>4	10 (1)
Outcome on adherence	
Adherent	174 (66.6)
Nonadherent	88 (33.4)

Evidence of Validity Based on Internal Structure

The unidimensional model obtained favorable fit indices (CFI=1; RMSEA=0, 90% CI 0-0.085; WRMR=0.072), as well as factor loadings around what was expected, except for item 2 (Do you always take your medication at the prescribed time? Table 3). After performing a second analysis without item 2, the results are adequate both at the level of fit indices (CFI=1; RMSEA=0, 90% CI 0-0.103; WRMR=0.047) and factor loadings (>0.50; Table 3).

Regarding reliability, the magnitudes were acceptable for both scores (Cronbach α=0.702) and construct (ω=0.718).

Finally, the correlation of adherence with viral load was statistically significant (r=0.128; P=.04), but not with recent CD4 T lymphocyte counts (r=0.015; P=.81).

On the relationship between undetectable viral load and adherence, those patients classified as adherent by the scale were 3 times more likely to be undetectable compared to those classified as nonadherent (OR 3.31, 95% CI 1.13-9.64, P=.04; Table 4).

XSL•FO
RenderX

Table 3. Factor loadings of the Simplified Medication Adherence Questionnaire scale.

Questions	First analysis	Second analysis
1. Do you ever forget to take your medication?	0.815	0.854
2. Do you always take your medication at the prescribed time?	–0.358	—[a]
3. Do you ever stop taking drugs if you feel sick or drink alcohol?	0.612	0.526
4. Did you forget to take your medication over the weekend?	0.383	0.374
5. In the last week, how many times did you not take a dose?	0.677	0.701

[a]Not applicable.

Table 4. Relationship between Simplified Medication Adherence Questionnaire adherence classification and viral load.

	Adherent	Nonadherent	OR[a] (95% CI)	*P* value
Undetectable viral load[b]	168	76	3.31 (1.13-9.64)	.04
Detectable viral load[c]	6	9	—[d]	—

[a]OR: odds ratio.

[b]Plasma viral load of ≤40 copies/mL.

[c]Plasma viral load >40 copies/mL.

[d]Not applicable.

Discussion

Principal Findings

Adherence is vital to achieve viral suppression and increase the quality of life of people living with HIV; however, its measurement represents a major challenge [26]. Although direct methods exist to measure adherence (eg, plasma drug concentration), they are costly and impractical for routine implementation in clinical settings with limited resources such as ours [8], thus, indirect measurement methods based on self-reporting may have a good performance and could be used routinely.

This study aimed to evaluate the reliability and validity of the SMAQ in people living with HIV/AIDS attending a public hospital in the western region of Mexico.

As for the factorial structure, 4 of the 5 original items were retained, and the one that was eliminated obtained the lowest factorial loading, and its response was oriented in the opposite direction to the other items, which reinforces the argument that this type of item usually presents methodological problems. This new four-item version obtained adequate fit and reliability indices (CFI=1; RMSEA=0, 90% CI 0-0.103; WRMR=0.047) [27].

The validity of its relationship with other variables was analyzed using an association with viral suppression (or undetectable viral load), finding a threefold greater probability of achieving viral suppression when classified as adherent (OR 3.31, 95% CI 1.13-9.64, *P*=.04).

The use of the SMAQ in the framework of explanatory studies could provide greater insight into the factors associated with adherence, providing evidence for the design of appropriate interventions to improve adherence in this patient population [28]. This, in turn, would strengthen the capacity of health services to enhance and promote quality care, ultimately yielding a positive long-term impact on people living with HIV.

Concerning previous studies, the study by Agala et al [10] also found adequate factor loadings, except for one item, like this work. However, there is no agreement on the eliminated item, which could be attributed to differences in the populations studied because the analysis of our study was only in men and theirs in women. This highlights the existing differences in adherence behaviors between these groups [10].

This situation underscores the need to consider sociodemographic and cultural characteristics when analyzing the evidence and validity of self-report tools such as the SMAQ.

Limitations

Regarding its limitations, it is important to recognize that, despite its simplicity, it may be subject to self-report bias [11]. Similarly, only men were considered due to the proportion that was recruited from both groups and to maintain statistical homogeneity [29]. This limitation does not allow for the opportunity to explore differences that could influence adherence between the two groups.

It is concluded that the SMAQ presents favorable evidence of validity per internal structure and association with viral suppression, as well as acceptable levels of reliability.

In future studies, we recommend that the SMAQ be analyzed at a psychometric level in the group of women to explore possible differences by gender. In addition, this tool will allow studies focused on adherence and its determinants to be carried out with a greater degree of precision.

Conclusions

The use of the SMAQ provides a valuable tool for assessing adherence among the Mexican population living with HIV. This evaluation is critical for enhancing health outcomes and

XSL•FO
RenderX

optimizing therapeutic interventions. Additionally, the SMAQ helps to strengthen public health programs by furnishing reliable data on treatment adherence, which can inform the development of targeted strategies to support patient engagement and medication management.

Acknowledgments

LEDMT contributed to the conception of the idea, the development of the protocol, analysis of results, and conclusions. LAGH worked on the development of the protocol, processing of results, and writing of this paper. JFAV did the writing of the discussion and the final review. PM-A, AVR, and VVRH selected and recruited participants. JAVR carried out the analysis of results and the development of the protocol. MGHG assisted in the general review of the document and the preparation of the final version. SD-L helped with the analysis of the scale, statistical analysis, and writing of the results.

Conflicts of Interest

None declared.

References

1. Hoja informativa—últimas estadísticas sobre el estado de la epidemia de sida. ONUSIDA. 2018. URL: https://www.unaids.org/es/resources/fact-sheet [accessed 2024-11-26]
2. Boletín de atención integral de personas con VIH. CENSIDA. 2024. URL: https://www.gob.mx/censida/articulos/boletin-de-diagnostico-y-tratamiento-antirretroviral-censida?idiom=es [accessed 2024-11-26]
3. Long-Term HIV treatment adherence for key populations. ONUSIDA. URL: https://www.fhi360.org/wp-content/uploads/drupal/documents/epic-long-term-hiv-adherence-guide.pdf [accessed 2024-11-26]
4. Oliveira RDS, Primeira MR, Santos WMD, Paula CCD, Padoin SMDM. Association between social support and adherence to anti-retroviral treatment in people living with HIV. Rev Gaucha Enferm 2020;41:e20190290 [FREE Full text] [doi: 10.1590/1983-1447.2020.20190290] [Medline: 32667420]
5. Tale AH, Tegegne AS, Belay DB. Predictors of viral load and medication adherence among HIV-positive adults under treatment at Felege-Hiwot Comprehensive Specialized Hospital, North-West, Ethiopia. HIV AIDS (Auckl) 2023;15:477-489 [FREE Full text] [doi: 10.2147/HIV.S422980] [Medline: 37593198]
6. Jorge D, Varela A, De PS, De P, Salud L, Nilza D. Sistema de vigilancia epidemiológica de vih informe histórico de VIH 1er trimestre. 2023. URL: https://www.gob.mx/salud [accessed 2024-11-26]
7. Adherence to Longterm Therapies. Geneva, Switzerland: World Health Organization; 2003.
8. López-Romero LA, Romero-Guevara SL, Parra DI, Rojas-Sánchez LZ. Adherencia al tratamiento: concepto y medición. Hacia la promoción de la salud 2016;21(1):117-137. [doi: 10.17151/hpsal.2016.21.1.10]
9. Knobel H, Alonso J, Casado J, Collazos J, González J, Ruiz I, GEEMA Study Group. Validation of a simplified medication adherence questionnaire in a large cohort of HIV-infected patients: the GEEMA Study. AIDS 2002;16(4):605-613. [doi: 10.1097/00002030-200203080-00012] [Medline: 11873004]
10. Agala CB, Fried BJ, Thomas JC, Reynolds HW, Lich KH, Whetten K, et al. Reliability, validity and measurement invariance of the simplified medication adherence questionnaire (SMAQ) among HIV-positive women in Ethiopia: a quasi-experimental study. BMC Public Health 2020;20(1):567 [FREE Full text] [doi: 10.1186/s12889-020-08585-w] [Medline: 32345253]
11. Ortega SF, Sánchez PJ, Pérez VM, Pereira PP, Muñoz CM, Lorenzo AD. Validación del cuestionario simplificado de adherencia a la medicación (SMAQ) en pacientes con trasplante renal en terapia con tacrolimus. Nefrologia 2011;31(6):690-696 [FREE Full text]
12. Merino-Soto C, Calderón-De La Cruz GA. Validity of peruvian studies on stress and burnout. Rev Peru Med Exp Salud Publica 2018;35(2):353-354 [FREE Full text] [doi: 10.17843/rpmesp.2018.352.3521] [Medline: 30183913]
13. Elangovan N, Sundaravel E. Method of preparing a document for survey instrument validation by experts. MethodsX 2021;8:101326 [FREE Full text] [doi: 10.1016/j.mex.2021.101326] [Medline: 34434840]
14. Leyva-Moral JM, Loayza-Enriquez BK, Palmieri PA, Guevara-Vasquez GM, Elias-Bravo UE, Edwards JE, et al. Adherence to antiretroviral therapy and the associated factors among people living with HIV/AIDS in Northern Peru: a cross-sectional study. AIDS Res Ther 2019;16(1):22 [FREE Full text] [doi: 10.1186/s12981-019-0238-y] [Medline: 31462291]
15. A-priori sample size calculator for structural equation models. Free Statistics Calculator. 2006. URL: https://www.danielsoper.com/statcalc/calculator.aspx?id=89 [accessed 2024-11-26]
16. Dominguez-Lara S. Propuesta de puntos de corte para cargas factoriales: una perspectiva de fiabilidad de constructo. Enfermería Clínica 2018;28(6):401-402. [doi: 10.1016/j.enfcli.2018.06.002]
17. Casanova CJ, Carmen RFM, Gómez GM. Manejo del paciente con VIH en la atención primaria. Actualización en Medicina Familiar. URL: https://www.comunidad.madrid/sites/default/files/doc/sanidad/manejo_paciente_vih_en_ap.pdf [accessed 2024-11-26]
18. McDonald RP, Ho MR. Principles and practice in reporting structural equation analyses. Psychol Methods 2002;7(1):64-82 [FREE Full text] [doi: 10.1037/1082-989x.7.1.64] [Medline: 11928891]

XSL•FO
RenderX

19. Muthén LK, Muthén BO. Statistical analysis with latent variables user's guide. 1998. URL: https://www.amazon.com/Mplus-Statistical-Analysis-Variables-Version/dp/B0050JZXQW [accessed 2024-11-26]
20. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res 1992;21(2):230-258. [doi: 10.1177/0049124192021002005]
21. DiStefano C, Liu J, Jiang N, Shi D. Examination of the weighted root mean square residual: evidence for trustworthiness? Struct Equation Model: Multidiscip J 2017;25(3):453-466. [doi: 10.1080/10705511.2017.1390394]
22. Ponterotto JG, Charter RA. Statistical extensions of Ponterotto and Ruckdeschel's (2007) reliability matrix for estimating the adequacy of internal consistency coefficients. Percept Mot Skills 2009;108(3):878-886. [doi: 10.2466/PMS.108.3.878-886] [Medline: 19725323]
23. Ferguson CJ. An effect size primer: a guide for clinicians and researchers. Prof Psychol Res Pr 2009;40(5):532-538. [doi: 10.1037/a0015808]
24. Epi InfoTM User Guide-Chapter 12-StatCalc 12. StatCalc. 2012. URL: https://www.cdc.gov/epiinfo/pdfs/userguide/12_statcalc.pdf [accessed 2024-11-26]
25. Jager J, Putnick DL, Bornstein MH. II. More than just convenient: the scientific merits of homogeneous convenience samples. Monogr Soc Res Child Dev 2017;82(2):13-30 [FREE Full text] [doi: 10.1111/mono.12296] [Medline: 28475254]
26. Nachega JB, Marconi VC, van Zyl GU, Gardner EM, Preiser W, Hong SY, et al. HIV treatment adherence, drug resistance, virologic failure: evolving concepts. Infect Disord Drug Targets 2011;11(2):167-174 [FREE Full text] [doi: 10.2174/187152611795589663] [Medline: 21406048]
27. Suárez-Álvarez J, Pedrosa I, Lozano L, García-Cueto E, Cuesta M, Muñiz J. Using reversed items in likert scales: a questionable practice. Psicothema 2018;30(2):149-158. [doi: 10.7334/psicothema2018.33] [Medline: 29694314]
28. Duthely LM, Sanchez-Covarrubias AP. Digitized HIV/AIDS treatment adherence interventions: a review of recent SMS/Texting mobile health applications and implications for theory and practice. Front Commun (Lausanne) 2020;5:530164 [FREE Full text] [doi: 10.3389/fcomm.2020.530164] [Medline: 33644162]
29. Girum T, Wasie A, Lentiro K, Muktar E, Shumbej T, Difer M, et al. Gender disparity in epidemiological trend of HIV/AIDS infection and treatment in Ethiopia. Arch Public Health 2018;76:51 [FREE Full text] [doi: 10.1186/s13690-018-0299-8] [Medline: 30237881]

Abbreviations

ART: antiretroviral treatment
CFI: comparative fit index
OR: odds ratio
RMSEA: root-mean-square error of approximation
SMAQ: Simplified Medication Adherence Questionnaire
WRMR: weight root-mean-square residual

Edited by T de Azevedo Cardoso; submitted 15.04.24; peer-reviewed by M Nacher, K Hoti; comments to author 11.10.24; revised version received 01.11.24; accepted 14.11.24; published 07.01.25.

Please cite as:
Del Moral Trinidad LE, González Hernández LA, Andrade Villanueva JF, Martínez-Ayala P, Valle Rodríguez A, Ruíz Herrera VV, Vizcaíno Résendiz JA, Herrera Godina MG, Dominguez-Lara S
Simplified Medication Adherence Questionnaire (SMAQ) for People Living With HIV in a National Hospital in Mexico: Instrument Validation Study
Interact J Med Res 2025;14:e59562
URL: https://www.i-jmr.org/2025/1/e59562
doi:10.2196/59562
PMID:39773472

©Luis Eduardo Del Moral Trinidad, Luz Alicia González Hernández, Jaime Federico Andrade Villanueva, Pedro Martínez-Ayala, Adriana Valle Rodríguez, Vida Veronica Ruíz Herrera, José Adán Vizcaíno Résendiz, Melva Guadalupe Herrera Godina, Sergio Dominguez-Lara. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 07.01.2025. 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

Interstep Variations of Stairways and Associations of High-Contrast Striping and Fall-Related Events: Observational Study

Sara A Harper[1,2], PhD; Chayston Brown[1], BS; Shandon L Poulsen[1], BS; Tyson S Barrett[3], PhD; Christopher J Dakin[1], PhD

[1]Department of Kinesiology and Health Science, Sorenson Legacy Foundation Center for Clinical Excellence, Utah State University, Logan, UT, United States

[2]Kinesiology Department, University of Alabama in Huntsville, Huntsville, AL, United States

[3]Highmark Health Research Institute, Pittsburgh, PA, United States

Corresponding Author:
Sara A Harper, PhD
Kinesiology Department
University of Alabama in Huntsville
301 Sparkman Drive NW
Huntsville, AL, 35899
United States
Phone: 1 256 824 2184
Email: sah0075@uah.edu

Abstract

Background: Interstep variations in step riser height and tread depth within a stairway could negatively impact safe stair negotiation by decreasing step riser height predictability and, consequently, increasing stair users' fall risk. Unfortunately, interstep variations in riser height and depth are common, particularly in older stairways, but its impact may be lessened by highlighting steps' edges using a high-contrast stripe on the top front edge of each step.

Objective: This study aimed to determine (1) if fall-related events are associated with greater interstep riser height and depth variations and (2) if such fall-related events are reduced in the presence of contrast-enhanced step edges compared with a control stairway.

Methods: Stair users were video recorded on 2 public stairways in a university building. One stairway had black vinyl stripes applied to the step's edges and black-and-white vertical stripes on the last and top steps' faces. The stairway with striping was counterbalanced, with the striped stairway than a control, and the control with stripes. Each stair user recorded was coded for whether they experienced a fall-related event. A total of 10,000 samples (observations) of 20 fall-related events were drawn with 0.25 probability from each condition to determine the probability of observing a distribution with the constraints outlined by the hypotheses by a computerized Monte Carlo simulation.

Results: In total, 11,137 individual stair user observations had 20 fall-related events. The flights that had 14 mm in interstep riser height variation and 38 mm in interstep depth variation were associated with 80% (16/20) of the fall-related events observed. Furthermore, 2 fall-related events were observed for low interstep variation with no striping, and 2 fall-related events were observed during low interstep variation with striping. A total of 20 fall-related events were observed, with 4 occurring on flights of stairs with low interstep variation. For stairs with high variability in step dimensions, 13 of 16 (81%) fall-related events occurred on the control stairway (no striping) compared with 3 of 16 (19%) on the high-contrast striping stairway. The distribution of fall-related events we observed between conditions likely did not occur by chance, with a probability of 0.04.

Conclusions: These data support the premise that a vision-based strategy (ie, striping) may counteract fall risk associated with interstep riser height and tread depth variation. Possibly, perception and action elicited through the horizontal-vertical illusion (striping) may have a positive impact on the incidence of fall-related events in the presence of high interstep riser height and depth variation. The findings of this study suggest that contrast enhancement (ie, striping) may be a simple and effective way to reduce the risk of falls associated with interstep variation, highlighting the potential for this approach to make a significant impact on fall prevention efforts.

(Interact J Med Res 2025;14:e60622) doi:10.2196/60622

XSL•FO
RenderX

KEYWORDS

stairs; stairway safety; riser height; tread depth; horizontal-vertical illusion; fall risk; fall prevention; videos; Monte Carlo simulation; public health; vision-based strategy; health promotion; adults; geriatric

Introduction

When approaching a stairway, stair users seemingly anticipate uniformity in the step riser height and tread depth [1,2]. However, this assumption may lead to fewer attentional resources being allocated to estimating these metrics, potentially compromising the safe negotiation of the stairway. Unfortunately, interstep riser height and tread depth variations are common. Often described as dimensional uniformity in building code, stair riser heights and tread depths shall be of uniform size and shape. The tolerance between the largest and smallest riser height, or tread depth, shall not exceed 3/8 inch in any flight of stairs. In our investigation, the range of dimension uniformity is referred to as interstep variation. Interstep variations, as small as 6 mm, between adjacent stair risers or treads can disrupt step cadence and increase the risk of accidents or falls [3]. One strategy to mitigate this risk involves applying a horizontal-vertical illusion and black stripes to the top front edge of each step, which could potentially decrease the frequency of slips, trips, and falls [4-6]. Recent research has demonstrated that adding a high-contrast stripe along the top front edge of each step [4,7-11] can lead to increases in heel clearance above the step. Similarly, adding vertical monochrome striping to the faces of the bottom and top steps can also enhance vertical foot clearance [4,5,12,13].

Although the exact mechanism behind these interventions is not fully understood, it is possible that they increase step height by drawing more attention to each step's edge [7] or by creating a horizontal-vertical illusion, which makes the steps appear taller than they actually are [14,15]. Ultimately, either mechanism may decrease the likelihood of a slip, trip, or fall on stairs with high interstep variation. We, therefore, hypothesized that when comparing 2 flights of stairs with similar interstep variability, the stairs with vertical monochrome striping and tread edge highlighters would record fewer fall-related events than stairs without this intervention. In addition, stairs with greater interstep variation in riser heights and tread depths would generally record more fall-related events than stairs with less interstep riser height and tread depth variability, but this effect would be lessened with the addition of monochrome striping and tread edge highlighters.

To test our hypotheses, we estimated the probability of observing a range of fall-related event distributions that could plausibly occur by chance, given our hypotheses [high or low interstep riser height and tread depth variations and control or striping intervention conditions]. We codified our hypotheses in a Monte Carlo simulation using 4 constraints based on our hypotheses. First, there would be more fall-related events on control flights of stairs (without the contrast intervention) with greater interstep variation than those with lower interstep variation. Second, both control and intervention (with the contrast enhancement) flights of stairs with low interstep variation should observe a comparable frequency of fall-related events (ie, a difference of less than 2 fall-related events between conditions). Third, fewer (less than or equal to half the number of) fall-related events should be observed with high interstep variation intervention stairs compared with the high interstep variation control stairs. Finally, the relative difference in fall-related events between high and low interstep variation stair flights in the control condition should be greater than on the intervention flights.

Methods

Participants

This cross-sectional study took place on 2 public stairways on the Utah State University campus. Video capture occurred on public stairways, and most stair users appeared to be young adults.

As previously discussed, 11,137 individual stair user observations were coded and balanced across the stairway conditions (control and intervention) and stairways (East and West) [16]. There were 7458 (66.97%) feminine observations and 3679 (33.03%) masculine observations, where most observations (n=10,970, 98.5%) were in the age group of 18-40 years. Additional participant details were described previously [6]. Given the observational nature of this study in a public space, no screening was performed in advance. Eligibility criteria were as follows: (1) inclusion—visually appearing 18 years of age and older, captured within local time (eg, 7 AM-5 PM) and (2) exclusion—use of assistive walking devices (eg, crutches and walking boots), individuals that did not transverse the stairs, or involved unusual stairway behavior, as described by Harper et al [6] were documented and removed during the data cleaning phase.

Protocol

High-resolution security cameras (8 megapixels, 4K Ultra HD, 3840×2160 resolution at 7 frames per second, Lorex cameras [Lorex Technology Inc]) were placed in the stairways to record stair users' behavior.

Intervention

High-contrast black vinyl film (Gerber High-Performance Series 220 vinyl film [Gerber Technology]) stripes (5.5 cm wide, 0.063-0.09 mm thick) were applied flush to the top front edge of each stair [4,9,12]. A total of 19 black-and-white vertical vinyl stripes (12 cycles/1 meter) were placed on the very bottom and top steps' faces [4,9,12].

Figure 1 [6] depicts the combined striping intervention. Stairway interstep riser height and depth variations were measured across every step, from the middle part of the step (Figure 2) [6].

XSL•FO
RenderX

Figure 1. Control and interventional stairway conditions. Intervention conditions are further illustrated to show step riser (face of step) and top of step (superior view) vinyl stripes. Step riser height=step height. Stairway width=step tread depth. Vinyl striping depth. Adapted from Harper et al [6], which is published under Creative Commons Attribution 4.0 International License [17].

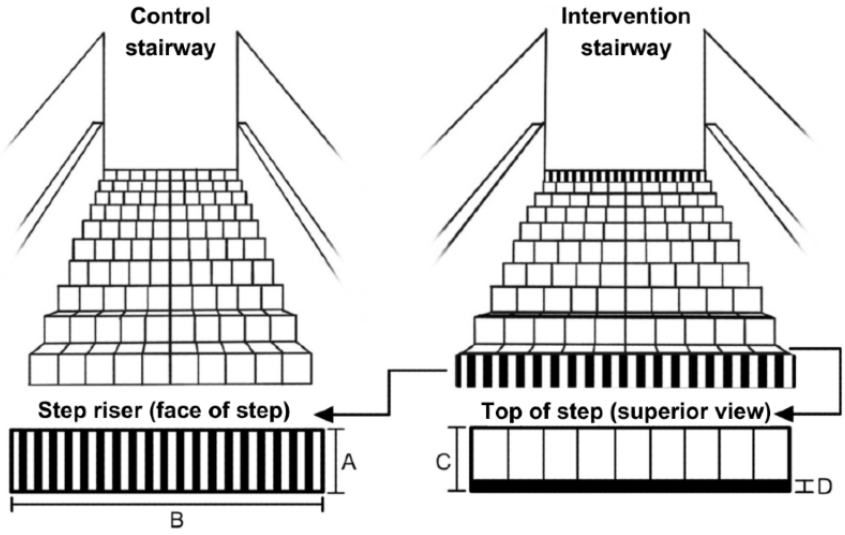
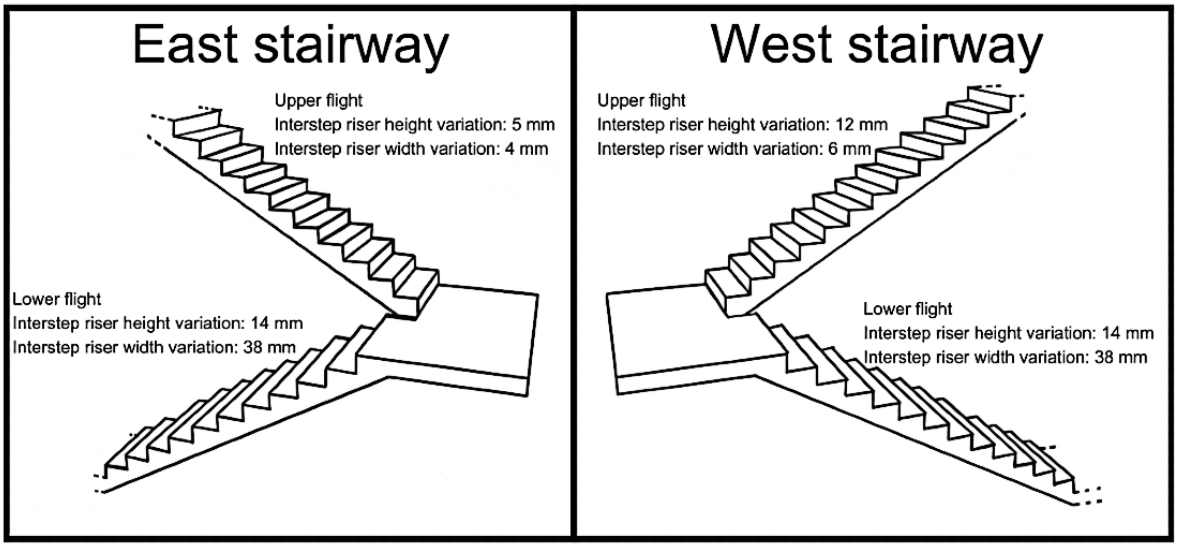


Figure 2. Real-world stairway design. East upper and lower stairway flights interstep riser height variation ranged between 5 mm and 14 mm, and interstep depth between 4 mm and 38 mm, respectively [6]. West upper and lower stairway flights' interstep riser height variation ranged between 12 mm and 14 mm, as well as interstep tread depth variation of 6 mm and 38 mm respectively, are shown [6]. Adapted from Harper et al [6], which is published under Creative Commons Attribution 4.0 International License [17].



In total, 48 steps were observed across 4 flights of stairs. The control stairway was unaltered and used to compare with the intervention stairway. Halfway through data collection, the intervention (striped) and control stairways were counterbalanced.

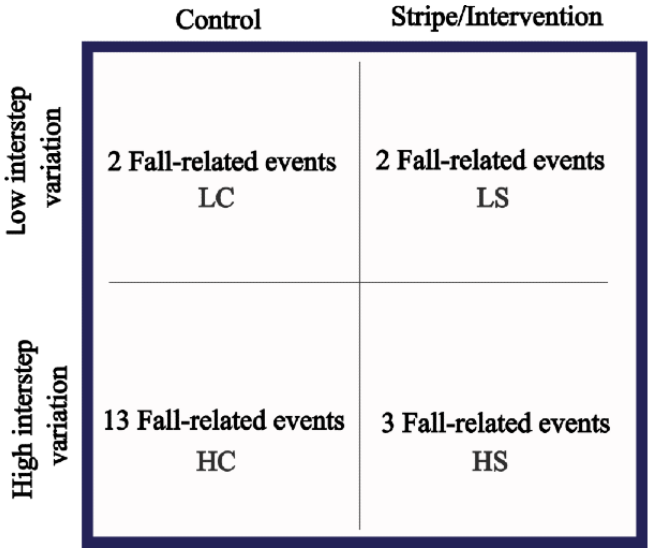
Measures

Each outcome variable was assessed by stairway location (East and West) and condition (intervention and control), as well as by stairway flight (lower and upper; Figure 3).

In total, 4 assumptions were used to code our hypotheses based on our a priori knowledge of the total number of fall-related events recorded (n=20): (1) the number of fall-related events high interstep variations, control condition>low interstep variations, control condition; (2) the difference between low interstep variations, stripe intervention and low interstep variations, control condition ≤2 fall-related events; (3) the number of falls in high interstep variations, control condition will be ≥2 times of high interstep variations stripe intervention; and (4) the difference between low interstep variations, control condition and high interstep variations, control condition > the difference between low interstep variations, stripe intervention and high interstep variations stripe intervention. The probability of a distribution meeting these assumptions occurring by chance is approximately 0.04 with a sample size of 20.

XSL•FO
RenderX

Figure 3. The probability of observing a range of distributions of fall-related events. The following assumptions evaluated the observed fall-related event distribution. HC: high interstep variations, control condition; HS: high interstep variations stripe intervention; LC: low interstep variations, control condition; LS: low interstep variations, stripe intervention.



Data Sources

We coded stair users' navigation direction (ascent and descent) and the presence of a fall-related event. As described previously, a fall-related event was coded if an observed stair user experienced a relatively subtle trip or slip, with minimal recovery action, through a complete loss in balance, resulting in a fall [6]. If a fall-related event occurred, the stairway and step number (starting from the bottom to the top) were recorded. To assess research bias or intercoder reliability, each week, one of the researchers would randomly select and evaluate 10% of the data collected for that week. If an error was present, a second coder reviewed all data recorded by the first coder for the week in question, and the second coder made a determination on the final coding record [6]. Data were encoded using Microsoft Excel.

Statistical Analysis

Data are presented as mean (SD) as well as count (percentage) of observed results. The distribution of fall-related events was assessed across stairway flights (interstep variations) and conditions (control and intervention) using a 0.5-inch threshold (approximately 13 mm), given that 75% (80/101) of stair accidents occurred in stairways with interstep riser height variations of ≥0.5 inch [18]. A Monte Carlo simulation in Julia [19] was used to estimate the probability of observing the distribution of fall-related events defined by our hypotheses by chance. Specifically, 10,000 samples (observations) of 20 fall-related events (refer to the Results section) were drawn with 0.25 probability from each condition to determine the probability of observing a distribution with the constraints outlined by the hypotheses through a computerized Monte Carlo simulation [20,21].

Ethical Considerations

Ethical approval was obtained from the institutional review board of Utah State University (10773). As an observational study, participants did not give written consent. Given the observational nature, participants were not compensated to participate. The video recordings were taken in a public setting, and only the approved research team had access to the identifiable data. Therefore, our video data are not available as supplementary material.

Results

Individuals who had any visible health-related conditions (eg, crutches and walking boots) were documented and removed before analysis. In addition, those who visually appeared under the age of 18 years were documented and removed before analysis. All steps were measured at the center of the steps. The average of the East and West stairways combined was 168 (SD 4) mm, with the average riser height of the West stairway steps being 171 (SD 3) mm, and the East stairway being 166 (SD 3) mm, independently. The average step tread depth for the East and West stairways was 328 (SD 8) mm, the average depth of the West stairway steps was 327 (SD 8) mm and the Easy stairway was 329 (SD 8) mm.

Of 20 observed fall-related events, 80% (n=16) of events were observed on the flights where interstep variations were the greatest (riser height ranged 14 mm and tread depth ranged 38 mm). In comparison, 4 of 20 fall-related events were observed on flights with lower interstep variation. Between East and West stairways, 7 of 20 (35%) fall-related events occurred on steps that had step riser heights greater than 1 SD from their mean including 2 fall-related events that occurred on the first step while ascending (step riser height=170 mm; flight mean 166, SD 3 mm), 3 events on the last step while descending the East, lower stairway (height=155 mm; mean 166, SD 3 mm), 1 event on the last step while ascending the East, upper stairway (height=170 mm; mean 166, SD 3 mm), and 1 event on the second-to-last step while ascending the West, lower stairway (height=173 mm; mean 171, SD 4 mm).

A total of 20 fall-related events were recorded. On flights where interstep riser height variation ranged 14 mm and step depth variation ranged 38 mm, 13 of 16 (81%) fall-related events occurred on the control stairway condition (no striping),

XSL•FO
RenderX

compared with 3 of 16 (19%) on the striped intervention stairway. Finally, 2 fall-related events were observed for low interstep variation with no striping, and 2 fall-related events were observed during low interstep variation with striping.

The estimated probability of observing data that fit the range of distributions constrained by our hypotheses and using a sample size of 20 fall-related events, was approximately 0.04 (Figure 3). This result suggests that (1) interstep variations may be contributing to falls and (2) adding a striping intervention to the stairs may reduce the impact of interstep variations on fall-related events.

Discussion

Principal Results

We sought to assess the impact of interstep riser height and tread depth variations on fall-related events (eg, slips, trips, or falls) occurring on stairways to determine if fall risk increases with greater interstep variations but is reduced by adding high-contrast striping. The lower flights of stairs, which had interstep riser height variations that ranged 14 mm and interstep depth variations that ranged 38 mm, accounted for 80% (16/20) of the observed fall-related events, supporting the notion that stairways with greater interstep variations may be associated with a greater risk of fall-related events. Furthermore, 35% (7/20) of fall-related events were observed on steps where interstep riser height variation was greater than 1 SD from the flight mean. Together, these results suggest that flights of stairs with greater interstep riser height variation exhibit more fall-related events than flights with lower interstep variation. When high-contrast striping was added to flights of stairs with high interstep variation, there were fewer fall-related events observed over a similar time period (3/20 fall-related events in the intervention vs 13/20 fall-related events in the control condition). This result suggests the addition of high-contrast striping to stair flights with high interstep variation may reduce the number of fall-related events resulting from interstep variation.

Comparison With Previous Work

Interstep Variations Increase Fall Risk

Interstep variations on stairways can have a profound negative impact on fall risk. Even minor interstep variations, such as a 6 mm variance in riser height [3], can disrupt step cadence and increase the likelihood of a fall, as can interstep variation in tread depth. Furthermore, a review of 80 stairway falls from 1992 to 2007 found that 60% (48/80) of riser height and 34% (27/80) of interstep depth variation were greater than 3/8 inch in a study by Cohen et al [7], and greater (3/8 inch compared with 0.5 inch) interstep variation that can disrupt cadence [3]. Francksen et al [22] found that adults could adjust their stepping behavior for increases in depth, but they could not adjust for interstep riser height variation over 10 mm. Stair-related fall injuries are also more commonly associated with interstep riser height variation than interstep tread depth variation [7]. Alternatively, the greater association of falls with interstep riser height variation could be due to the increased frequency of observing interstep riser heights compared with tread depth variations (eg, > 3/8 inch is observed more often in riser heights [48/80, 60%] than tread depths [27/80, 34%]) [7]. Nevertheless, additional research is needed to explore the extent and impact of interstep variation on stair-related falls. In this study, we observed that 80% (16/20) of fall-related events on stairs occurred when the interstep riser height variation was 14 mm and depth was 38 mm. Since interstep riser height and depth variations were both present in our observational design, we are unable to distinguish which of these factors may have had a greater role.

The Intervention Was Associated With Reduced Fall-Related Events With Interstep Variation

Broadly, our findings indicate that greater interstep variations are associated with an increased frequency of fall-related events. Importantly, the frequency of these fall-related events decreased when the intervention was present. Our data suggest that adding high-contrast stripes may reduce fall-related events on stairs when large interstep variations are present. Of the 20 fall-related events observed, 16 were on the flights of stairs with larger interstep variation, but only 19% (3/16) occurred when the intervention was present, whereas 81% (13/16) occurred when the intervention was absent.

Mechanisms Contributing to the Reduction in Fall-Related Events

The intervention may induce a horizontal-vertical illusion by the intervention. This occurs when horizontal and vertical lines of similar length are presented together (like the letter “T”), which results in an illusory sense that the vertical line is longer than the horizontal one. The intervention used in this study included black vinyl stripes applied to each step’s top front edge and black-and-white vertical stripes on the face of the first and last steps. Together, the abutting edges of the combined striping that formed “T”-like configurations could have contributed to an increased perceived step riser height [14], resulting in greater step heights. Indeed, previous research suggests that under similar experimental conditions, perceived step height is increased and is associated with an increase in the height of the step taken [12,13,15]. In further support for this mechanism, the horizontal-vertical illusion effect is reduced when only edge highlighters are present [13], suggesting that the vertical lines contribute to the increased step height and, perhaps, that the horizontal-vertical illusion is a primary mechanism contributing to the reduction in fall-related events we observed with the intervention. It is also possible that other intervention formats, such as those that could induce the Müller-Lyer illusion [23], could reduce fall-related events. The Müller-Lyer illusion occurs when the perceived length of a line is influenced by the orientation of arrow-like segments attached to its ends. Lines of the same length appear shorter or longer depending on whether the arrowheads at the ends point inward or outward. By manipulating the direction of the fins at the end of the lines, perhaps step height could be increased or decreased depending on the particular interstep variation. Recent outdoor observational research suggests that greater vertical foot clearance occurs when a “fins out” configuration is applied to a 2-step stairway [24], which is the illusion’s expected effect.

XSL•FO
RenderX

Future research could examine whether a "fins in" configuration reduces step height.

By accentuating the steps' edges, the intervention may draw attention to them, thereby enhancing the accuracy with which stair users can estimate the dimensions of each step and then compensate for irregularities. Future research could investigate whether the presence of striping increases awareness of interstep variation, thereby evaluating this potential mechanism. In addition, the novelty of the intervention itself may have drawn attention to the stairs. Schomaker and Meeter [25] suggest that novel visual stimuli, such as a change in contrast, may increase attention to the stairs [26]. It is worth noting that the striping intervention was installed several days before data collection began. It is likely that some stair users had exposure to the striping before the start of video recording and this early exposure could have reduced the novelty of the striping effect before data collection, thus reducing the drawing effect. We also expect that this effect would have worn off over time as many of the observed stair users traverse the stairs frequently due to regular classes in the building.

What Should Be Done About Excessive Interstep Variations

Maybe the most important challenge associated with observing increased fall-related events with greater interstep variations is what to do about it. Is the risk of falls sufficient to warrant widespread evaluation and enforcement of building codes? Although we do not provide recommendations here, conducting assessments, enforcing regulations during construction, and evaluating older stairways (likely to exhibit the greatest interstep variation) may reduce fall-related events, especially when considering cost-effectiveness. For older stairways, as used in this study (built in 1971, and at the time of construction, the 1967 Uniform Building Code [Sec 3305 (d)] that was in place in the United States required the maximum interstep variation in riser height and tread depth to be no more than 3/16 inch), adding an intervention like painting stripes might be the most cost-effective way (an estimated US $288 for the high-contrast striping used here) to reduce the impact of interstep variations (assuming the application or materials used to apply the striping do not themselves increase fall risk through reduced or increased friction, or materials peeling). While increased interstep variation is associated with a greater risk of fall-related events, there are cost-effective interventions that can help reduce this risk. By enforcing building codes and evaluating stairways for interstep variation, we can work toward creating safer environments for everyone.

Limitations

While these results are promising, we acknowledge several limitations of this study. Even with over 10,000 observations, we only observed 20 fall-related events. A larger sample of falls would provide a more precise estimate of the differences between flights of stairs and strengthen the inferences that could be drawn from these data. Since most of the observations in this study were younger adults, future work should consider targeting older or clinical populations (eg, those with visual impairment and mobility-related limitations) to determine if such an intervention could reduce fall-related events. However, future designs will need to consider comparing historical fall frequency records to future intervention fall frequencies rather than using a control condition if it could pose a fall risk to these populations. Furthermore, we did not include Cohen κ, as a measure of interrater reliability. In addition, the lack of a validated questionnaire, such as Yang et al [27], are methodical limitations and should be included in future studies. In addition, given the emphasis on younger adults in this study, it is unknown whether the striping intervention's impact is greater in younger versus older adults. Finally, since the steps used here had interstep variations in both riser heights and tread depths, future observational designs could assess these 2 factors independently to determine the impact of each on fall risk.

Future Directions

Given the considerable negative impact of falls on public health, continued research is necessary to improve safety on stairways. In addition, programs, such as educational campaigns, could be used to raise awareness of factors that contribute to falls [28] and perhaps to help motivate small actions, such as painting stripes on problematic stairways that could have a big impact on public health and provides support for scaling up effective public health interventions for long-term population health benefits.

Conclusions

This study highlights the importance of addressing interstep variations in stairways to reduce the risk of fall-related events. By understanding the factors that contribute to fall risk and implementing cost-effective interventions, we can work toward creating safer environments for everyone. The findings of this study suggest that contrast enhancement (ie, striping) may be a simple and effective way to reduce the risk of falls associated with interstep variations, highlighting the potential for this approach to make a significant impact on fall prevention efforts.

Acknowledgments

We would like to thank Christopher Long, Samantha Corbridge, Alex Braeger, Brevin J. Zollinger, Kenneth Harrison, Travis Boman, Amy Hale, Emmalee Rolfe, McKay Wilding, Erika Larson for their assistance in data coding.

This work was supported by the Utah State University Undergraduate Research and Creative Opportunity Award (2020).

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available since the video data are not deidentified.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Archea J, Collins BL, Stahl FI. Guidelines for stair safety, Building Science Series, National Institute of Standards and Technology, Gaithersburg, MD; 1979. 1979 May. URL: https://doi.org/10.6028/Nbs.bss.120 [accessed 2024-10-03]
2. Nemire K, Johnson DA, Vidal K. The science behind codes and standards for safe walkways: changes in level, stairways, stair handrails and slip resistance. Appl Ergon 2016 Jan;52:309-316. [doi: 10.1016/j.apergo.2015.07.021] [Medline: 26360223]
3. Templer J. The Staircase. Second Edition. Cambridge: Massachusetts Institute of Technology; 1992.
4. Elliott DB, Foster RJ, Whitaker D, Scally AJ, Buckley JG. Analysis of lower limb movement to determine the effect of manipulating the appearance of stairs to improve safety: A linked series of laboratory-based, repeated measures studies. In: Public Health Research. Southampton (UK): NIHR Journals Library; 2015.
5. Elliott DB, Vale A, Whitaker D, Buckley JG. Does my step look big in this? A visual illusion leads to safer stepping behaviour. PLoS One 2009;4(2):e4577 [FREE Full text] [doi: 10.1371/journal.pone.0004577] [Medline: 19240790]
6. Harper SA, Long C, Corbridge S, Barrett TS, Braeger A, Zollinger BJ, et al. Stairway visual contrast enhancement to reduce fall related events. Journal of Ergonomics 2022;12(303). [doi: 10.35248/2165-7556-22.12.303]
7. Cohen J, LaRue CA, Cohen HH. Stairway falls an ergonomics analysis of 80 cases. Professional Safety 2009;54(1) [FREE Full text]
8. den Brinker BPLM, Burgman LJ, Hogervorst SMJ, Reehorst SE, Kromhout S, van der Windt J. The effect of high-contrast marking of treads on the descent of stairways by low-vision people. International Congress Series 2005 Sep;1282(9):502-506. [doi: 10.1016/j.ics.2005.05.165]
9. Foster RJ, Hotchkiss J, Buckley JG, Elliott DB. Safety on stairs: influence of a tread edge highlighter and its position. Exp Gerontol 2014 Jul;55:152-158 [FREE Full text] [doi: 10.1016/j.exger.2014.04.009] [Medline: 24768822]
10. Thomas NM, Skervin TK, Foster RJ, Parr JV, Carpenter MG, O'Brien TD, et al. Influence of step-surface visual properties on confidence, anxiety, dynamic stability, and gaze behaviour in young and older adults. Hum Mov Sci 2021 Jun;77:102774. [doi: 10.1016/j.humov.2021.102774] [Medline: 33676032]
11. Zietz D, Johannsen L, Hollands M. Stepping characteristics and Centre of Mass control during stair descent: Effects of age, fall risk and visual factors. Gait Posture 2011 Jun;34(2):279-284. [doi: 10.1016/j.gaitpost.2011.05.017] [Medline: 21680187]
12. Foster RJ, Whitaker D, Scally AJ, Buckley JG, Elliott DB. What you see is what you step: the horizontal-vertical illusion increases toe clearance in older adults during stair ascent. Invest Ophthalmol Vis Sci 2015 May;56(5):2950-2957. [doi: 10.1167/iovs.14-16018] [Medline: 26024081]
13. Skervin TK, Thomas NM, Schofield AJ, Hollands MA, Maganaris CN, Foster RJ. The next step in optimising the stair horizontal-vertical illusion: Does a perception-action link exist in older adults? Experimental Gerontology 2021 Jul;149:111309. [doi: 10.1016/j.exger.2021.111309] [Medline: 33716111]
14. Schofield AJ. Second-order texture gratings produce overestimation of height in depictions of rectangles and steps. Vision Res 2022 Nov;200:108101 [FREE Full text] [doi: 10.1016/j.visres.2022.108101] [Medline: 35908371]
15. Skervin TK, Thomas NM, Schofield AJ, Hollands MA, Maganaris CN, O'Brien TD, et al. Using a stair horizontal-vertical illusion to increase foot clearance over an inconsistently taller stair-riser. PLoS One 2021;16(9):e0257159 [FREE Full text] [doi: 10.1371/journal.pone.0257159] [Medline: 34520496]
16. Brown CB, Barrett TS, Long C, Corbridge S, Braeger A, Zollinger BJ, et al. Step edge highlighters and illuminance changes influence stair descent in a real-world setting. Ergonomics 2022 Nov 04;66(9):1219-1228. [doi: 10.1080/00140139.2022.2141347] [Medline: 36314061]
17. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2025-01-03]
18. Miller JA, Esmay ML. Nature and causes of stairway fall. 1961. URL: https://elibrary.asabe.org/abstract.asp?aid=41024&t=2&redir=&redirType= [accessed 2024-12-03]
19. Bezanson J, Edelman A, Karpinski S, Shah VB. Julia: A Fresh Approach to Numerical Computing. SIAM Rev 2017 Jan;59(1):65-98. [doi: 10.1137/141000671]
20. Kroese DP, Brereton T, Taimre T, Botev ZI. Why the Monte Carlo method is so important today. WIREs Computational Stats 2014 Jun 20;6(6):386-392. [doi: 10.1002/wics.1314]
21. Mundform DJ, Schaffer J, Kim MJ, Shaw D, Thongteeraparp A, Supawan P. Number of Replications Required in Monte Carlo Simulation Studies: A Synthesis of Four Studies. J. Mod. App. Stat. Meth 2011 May 01;10(1):19-28 [FREE Full text] [doi: 10.22237/jmasm/1304222580]
22. Francksen NC, Ackermans TMA, Holzer D, Ebner SA, Maganaris CN, Hollands MA, et al. Negotiating stairs with an inconsistent riser: Implications for stepping safety. Appl Ergon 2020 Sep;87:103131. [doi: 10.1016/j.apergo.2020.103131] [Medline: 32501252]

XSL•FO
RenderX

23. Müller-Lyer FC. Optische urteilstauschungen. Archiv fur anatomie und physiologie. Physiologische Abteilung 1889;2:263-270 [FREE Full text]
24. Shim J, Sims B, Hudnall JT, Rylander J. The effects of Müller-Lyer illusion on toe clearance height in stair ascent. Gait & Posture 2022 Oct;98:78-84. [doi: 10.1016/j.gaitpost.2022.08.019] [Medline: 36063571]
25. Schomaker J, Meeter M. Novelty Enhances Visual Perception. PLoS ONE 2012 Dec 5;7(12):e50599 [FREE Full text] [doi: 10.1371/journal.pone.0050599] [Medline: 23227190]
26. Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. PLoS Biol 2015 Apr;13(4):e1002128 [FREE Full text] [doi: 10.1371/journal.pbio.1002128] [Medline: 25901488]
27. Yang Y, Feldman F, Leung PM, Scott V, Robinovitch SN. Agreement between video footage and fall incident reports on the circumstances of falls in long-term care. J Am Med Dir Assoc 2015 May 01;16(5):388-394. [doi: 10.1016/j.jamda.2014.12.003] [Medline: 25669670]
28. Harper SA, Corbridge S, Long C, Barrett TS, Braeger A, Zollinger BJ, et al. Safe stairway negotiation: Role of distractions and handrail use. J Safety Res 2022 Sep;82:314-322. [doi: 10.1016/j.jsr.2022.06.007] [Medline: 36031259]

Edited by T de Azevedo Cardoso; submitted 16.05.24; peer-reviewed by S Rietdyk, A Gebre; comments to author 17.09.24; revised version received 28.09.24; accepted 21.11.24; published 08.01.25.

Please cite as:
Harper SA, Brown C, Poulsen SL, Barrett TS, Dakin CJ
Interstep Variations of Stairways and Associations of High-Contrast Striping and Fall-Related Events: Observational Study
Interact J Med Res 2025;14:e60622
URL: https://www.i-jmr.org/2025/1/e60622
doi:10.2196/60622
PMID:

©Sara A Harper, Chayston Brown, Shandon L Poulsen, Tyson S Barrett, Christopher J Dakin. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 08.01.2025. 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

Understanding Loneliness Through Analysis of Twitter and Reddit Data: Comparative Study

Hurmat Ali Shah[1], PhD; Mowafa Househ[1], PhD

Hamad Bin Khalifa University, Doha, Qatar

Corresponding Author:
Hurmat Ali Shah, PhD
Hamad Bin Khalifa University
Education City
Doha
Qatar
Phone: 974 70620740 ext 00
Email: hshah@hbku.edu.qa

Abstract

Background: Loneliness is a global public health issue contributing to a variety of mental and physical health issues. It increases the risk of life-threatening conditions and contributes to the burden on the economy in terms of the number of productive days lost. Loneliness is a highly varied concept, which is associated with multiple factors.

Objective: This study aimed to understand loneliness through a comparative analysis of loneliness data on Twitter and Reddit, which are popular social media platforms. These platforms differ in terms of their use, as Twitter allows only short posts, while Reddit allows long posts in a forum setting.

Methods: We collected global data on loneliness in October 2022. Twitter posts containing the words “lonely,” “loneliness,” “alone,” “solitude,” and “isolation” were collected. Reddit posts were extracted in March 2023. Using natural language processing techniques (valence aware dictionary for sentiment reasoning [VADER] tool from the natural language toolkit [NLTK]), the study identified and extracted relevant keywords and phrases related to loneliness from user-generated content on both platforms. The study used both sentiment analysis and the number of occurrences of a topic. Quantitative analysis was performed to determine the number of occurrences of a topic in tweets and posts, and overall meaningful topics were reported under a category.

Results: The extracted data were subjected to comparative analysis to identify common themes and trends related to loneliness across Twitter and Reddit. A total of 100,000 collected tweets and 10,000 unique Reddit posts, including comments, were analyzed. The results of the study revealed the relationships of various social, political, and personal-emotional themes with the expression of loneliness on social media. Both platforms showed similar patterns in terms of themes and categories of discussion in conjunction with loneliness-related content. Both Reddit and Twitter addressed loneliness, but they differed in terms of focus. Reddit discussions were predominantly centered on personal-emotional themes, with a higher occurrence of these topics. Twitter, while still emphasizing personal-emotional themes, included a broader range of categories. Both platforms aligned with psychological linguistic features related to the self-expression of mental health issues. The key difference was in the range of topics, with Twitter having a wider variety of topics and Reddit having more focus on personal-emotional aspects.

Conclusions: Reddit posts provide detailed insights into data about the expression of loneliness, although at the cost of the diversity of themes and categories, which can be inferred from the data. These insights can guide future research using social media data to understand loneliness. The findings provide the basis for further comparative investigation of the expression of loneliness on different social media platforms and online platforms.

(Interact J Med Res 2025;14:e49464) doi:10.2196/49464

KEYWORDS

health informatics; loneliness informatics; loneliness theory; health effects; loneliness interventions; social media; lonely; loneliness; isolation; mental health; natural language processing; tweet; tweets; comparative analysis

XSL•FO
RenderX

Introduction

Loneliness not only affects quality of life but also leads to other mental health issues, thus burdening the public health service system. The monetary loss as a result of loneliness is estimated to be between US $8074.80 and US $12,077.70 per person per year in the United Kingdom [1]. The monetary cost of lost days and loss in productivity is estimated to be US $3.14 billion per year for employees in the United Kingdom. Loneliness is shown to be associated with a high risk for multiple health conditions such as physical and mental health issues, dementia, and early mortality [2].

Loneliness is formally defined as, "the unpleasant experience that occurs when a person's network of social relationships is deficient in some important way, quantitatively or qualitatively" [3]. Loneliness, thus, is the perceived and subjective dissonance between one's desired and actual social contacts and relationships. It is difficult to know whether a person is lonely when there is no direct reporting by the person, as loneliness is a very subjective phenomenon. As opposed to social isolation, which is objective and points to a lack of any social connection, loneliness is hard to identify. Loneliness also follows a U-shaped curve demographically, that is, young and old people are the loneliest [4]. Because social media and technology are more frequently used by the younger generation, they are helpful for analyzing loneliness. A previous report provided a theoretical explication of loneliness, loneliness informatics, and the gaps in research on the topic of loneliness informatics, that is, the application of informatics tools to study loneliness [5].

The causes of mental health issues can vary from genetic factors to social and economic factors, and might involve immediate family or finding meaning in life. The result can be withdrawal from human bonds and touch. Loneliness can be addressed through different interventions. However, there is a need to understand the prevalence of loneliness to devise technology-based and community-oriented strategies. Technology may have resulted in fragmented and individualized existence, but it can also be a great healer. The rise of social media has transformed the way we interact with others, offering new opportunities for social connection and communication. Loneliness is a common experience that can have negative effects on mental and physical health, and social media use has been implicated as a potential contributor to loneliness.

To better understand loneliness, this study performed a comparative analysis of Twitter and Reddit data. The aim of this study was to analyze data from these 2 different social media platforms to identify the topics and themes highly associated with the mention of loneliness. We aimed to identify the associations of socioeconomic, political, and personal-psychological factors with the feeling and expression of loneliness. We were not interested in finding out how many people are lonely. We wanted to assess what kind of expression is associated with loneliness. Both platforms are popular social media sites that allow users to post and interact with others, but they differ in terms of their user base and content focus. Twitter is a micro-blogging site with a broad user base that covers a wide range of topics, while Reddit is a forum-based site with niche communities focused on specific topics.

Loneliness is a subjective experience. Social media has become one of the best sources to study loneliness because loneliness must be studied by self-reporting of this subjective feeling. Understanding the nature of social media platforms viz-a-viz the kind of expression about loneliness is important. This study is a step toward understanding the difference in how loneliness is expressed via different social media platforms. The aim is not to conceptualize loneliness per se as this is beyond the scope of this study, but our analysis will provide insights into what terms, topics, and categories occur more frequently in the mention of loneliness.

The aim of the study is not to discuss the qualitative difference between Reddit and Twitter. The difference between both platforms will stand for all topics of analysis. We do not claim to present the differences between the 2 platforms, which is beyond the scope of this paper. Moreover, discussing the differences between different social media platforms may not be of research interest. The research interest is to assess whether the expression of loneliness varies on different social media platforms. We found that there was some variance. The actual reasons may not be known as we are not aware of any research that mentions why people prefer one social media platform over another. Using natural language processing techniques, we analyzed user-generated content on both platforms to identify and extract keywords and phrases related to loneliness. We then compared the prevalence and nature of loneliness-related content on Twitter and Reddit to identify common themes and trends related to loneliness across the 2 platforms. We wanted to understand the topics mostly associated with loneliness and their frequency of occurrence with the mention of loneliness. Before carrying out this analysis, sentiment analysis of tweets was performed through the valence aware dictionary for sentiment reasoning (VADER) tool from the natural language toolkit (NLTK) [6]. The research questions were as follows:

1. What topics and themes can be identified from the data collected on the expression of loneliness from social media? It is important to understand what the data covers and then to carry out an objective comparison.
2. Are there differences in the expression of loneliness on different social media platforms?
3. How the topics and themes associated with loneliness across social media platforms can be divided into various categories?

Social media data with public access are available for Twitter and Reddit. Facebook, a popular social media platform, does not provide access to information based on key terms. As of October 2023, Facebook provides access to an Ad Targeting dataset. Quora, which is a blog-based social media platform, does not have any public application programming interface (API) to provide access to its information. Reddit and Twitter were selected as their data were publicly available when we started collecting data.

Reddit and Twitter as well as other social media platforms can have their own audience, and it cannot be said with any certainty why people use or prefer one platform over the other. This paper

XSL•FO
RenderX

does not provide an answer to this question. The characteristics of each social media platform are broadly true for any kind of data extracted from them. In this study, we wanted to investigate whether the expression of loneliness varies on different social media platforms. The objective reasons or causal reasons for the differences are beyond the scope of this study.

Several systematic and scoping reviews have been performed to better understand loneliness. Understanding loneliness theoretically and evaluating its relation to mental health have been the topics of several studies [7]. Theoretical studies on loneliness have established the negative effects of loneliness. Similarly, from the health informatics perspective, there have been multiple systematic reviews of the connection between loneliness and mental health [8]. Moreover, there have been studies dealing with the application of technology-based interventions to cope with loneliness [5].

A comparative methodology has been used to compare countries in terms of the parameters of public health [9,10]. Although this study does not involve a country-wise comparative analysis, it is pertinent to note that a comparative analysis of different topics is an established practice. Foufi et al [11] used Reddit as a source to mine textual information about chronic diseases, the entities for the expression of diseases and conditions, and their relations. Similarly, Reddit was used by Schrading et al [12] for an analysis of domestic abuse. They developed a classifier based on the data available on Reddit to classify whether a post is about domestic abuse. Twitter data were analyzed by Tsai and Wang [13] to evaluate people's attitudes toward public health policies. The use of Twitter to analyze mental health or other public health conditions during COVID-19 is a prevalent approach, with studies focusing on a range of issues from insights into the pandemic [14] to analyzing social network data [15] and discerning emotions about COVID-19 [16].

Comparison of social media data for different purposes has been carried out in the literature because each social media platform is associated with a particular kind of user and a particular mode of expression. Cinelli et al [17] used Gab, Facebook, Reddit, and Twitter to assess the difference of opinion across different platforms. Curiskis et al [18] focused specifically on Twitter and Reddit for document clustering and topic modeling, while Gozzi et al [19] performed a comparative analysis of media coverage on Reddit and Wikipedia. Similarly, both Twitter and Reddit were used to monitor the spread of hateful content in the wake of COVID-19 [20].

This study builds on literature reviews involving a comparative analysis of social media and other such comparative techniques to better understand loneliness. This study attempts to assess how loneliness is expressed on Reddit and Twitter. The reason for this analysis is to clarify the roles of these platforms in research on loneliness. As social media can provide an intimate perspective on the experience and expression of the feeling of loneliness, understanding how different social media can contribute to research on loneliness is important. Loneliness is a subjective experience, and analysis of social media data can play a fundamental role in informing strategies to counter loneliness and understand its dynamics. This study attempts to clarify how different social media platforms can be used to understand loneliness. The findings of this study will be helpful in devising a large framework using social media for understanding loneliness.

Methods

Ethical Considerations

This study was exempt from seeking ethical approval as it analyzed publicly available online content. The study did not use the identity of the people from the data but provided an overall picture based on opinions expressed publicly. Data were collected from the public domain, and the data analysis involved aggregate terms, which were not concerned with individual statements. There were no relevant guidelines from the institutional review boards of the authors' institutions regarding the use of publicly available social media data. In the literature, no ethical approval has been taken for using social media data to perform different assessments.

Methodology

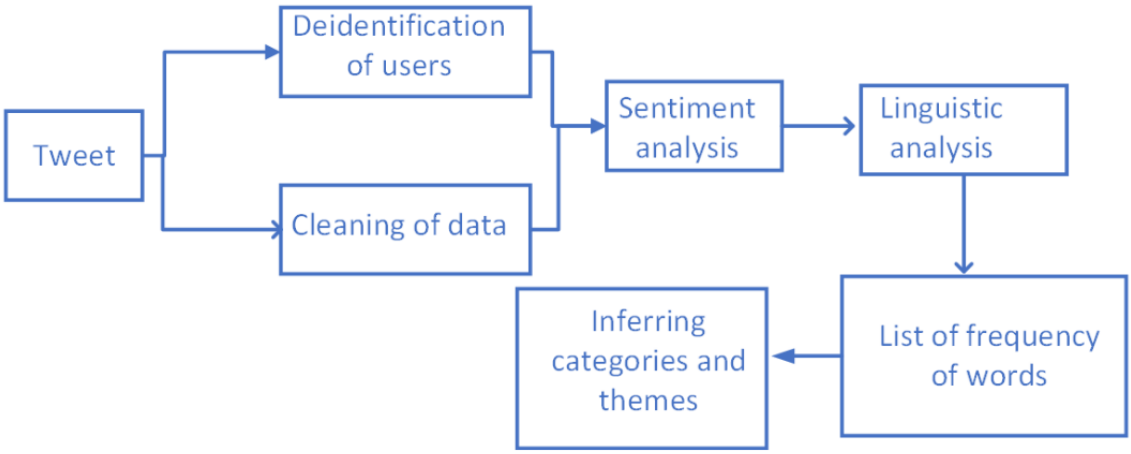
We used social intelligence analysis (SIA) to identify the themes associated with the expression of loneliness. SIA is a broad theme that incorporates multiple social media sources such as Facebook, Reddit, and Quora. SIA is important to gain insights into users' data and, in our case, understand the dynamics of loneliness. While SIA can be used for a variety of purposes, such as mining content to create stories and finding trends, we used SIA for sentiment analysis of collected data on loneliness. This is a comparative study to assess the themes associated with loneliness across different social media platforms. Therefore, we used Twitter and Reddit to collect data. The following subsections describe in detail the processes of data collection and analysis for Twitter and Reddit data.

Twitter

We performed an analysis of publicly available data of users posting about loneliness. Figure 1 presents our pipeline of the analysis of data collected from Twitter. Twitter is a social media platform that is used for connectivity and opinion sharing, and allows users to post via short messages consisting of 280 characters. Twitter provides access to user data through its publicly available Twitter API for developers. Relevant tweets about loneliness were gathered and stored in a database. Twitter data analysis involved the following 3 stages: (1) data collection (tweet collection); (2) division of the collected data into negative and other tweets through preliminary analysis (sentiment analysis of tweets); and (3) further analysis of tweets with negative sentiments through manual coding to find relevant themes and categories (manual coding and analysis of tweets).

XSL•FO
RenderX

Figure 1. Pipeline for processing Twitter data.

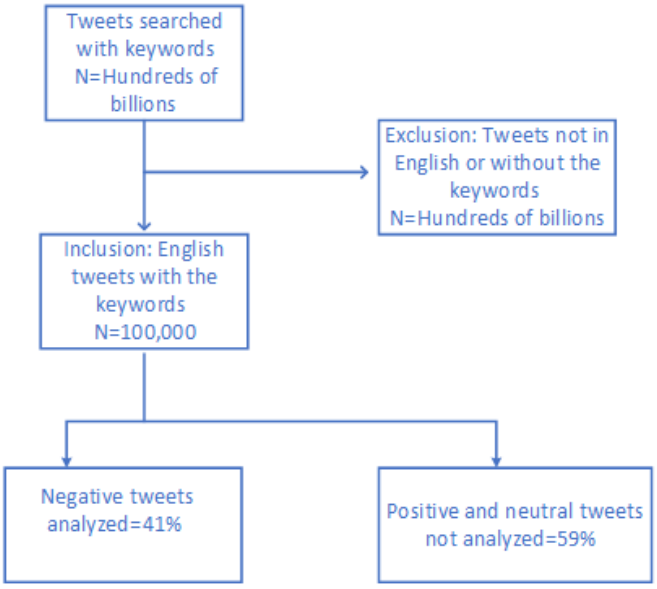


Tweet Collection

Tweets containing the keywords "lonely," "loneliness," "alone," "isolated," and "isolation" were collected. In theoretical literature, the words "loneliness" and "lonely" are used to describe the feeling under consideration in this study. The authors of a previous report [21] collected Twitter data based on the keywords "lonely" and "alone." We went further and included the synonyms and related words of loneliness for collecting our Twitter data.

We did not want to exhaustively search for data from a specific country because we wanted the collected data to be a proof of concept. We focused exhaustively on cities or countries and collected more data. The majority of tweets were from the United States (38%) and India (24%). The rest were from various countries across the globe. Tweets were extracted from these 2 countries to make a subdataset. This was meant to reflect the majority composition of the dataset. The subdataset contained 100,000 tweets. The collected data were merged based on location, user ID, and tweet ID to identify tweets belonging to the United States and India. We did not have control over the countries from which the data were collected. However, after data collection, we found that most tweets were from the United States and India. From the data collected, we made another dataset that contained 60% data from the United States and 40% from India. No city-wise allocation of data or analysis was carried out. Figure 2 presents the process of collecting data from Twitter and the process of tweet analysis.

Figure 2. STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) diagram for Twitter data.



Sentiment Analysis of Tweets

The next step was to perform a preliminary analysis of the collected tweets through sentiment analysis. This step is required because loneliness is a negative feeling. To determine how it is expressed in its meaning for mental health analysis, we filtered out tweets where the expression of loneliness was negative. If we decided to report all tweets that contained feelings of loneliness, we would not have required a further step. In our case, the problem becomes determining the association of themes (which may represent loneliness) with the keywords depicting loneliness. For instance, we had to assess the relationship of "hurt," "sick," "tired," "sleep," etc with the expression of loneliness. This task is usually carried out through the association of lexicon categories with tweets including the word "lonely" or "alone." The collected tweets also contained

XSL•FO
RenderX

metaphorical use of lonely or loneliness, which does not pertain to our use of loneliness. Such mentions of loneliness are represented by positive and neutral sentiment tweets. The definition of loneliness in this paper connotates a negative feeling. While loneliness can also be a positive or neutral feeling for some people or some situations, when it comes to its association with mental health issues, the negative consequences of loneliness must be considered.

Therefore, to perform the task of separating tweets with negative connotations for loneliness, we used sentiment analysis. Sentiment analysis was carried out after cleaning the data, such as removing redundant characters, numbers, special characters, user profile IDs, and information like "retweets." For finding tweets with negative sentiment, we used VADER based on Python's NLTK. VADER is suited for microblog content, such as that of Twitter. VADER combines the lexicon, that is, dictionary-based analysis, and rule-based approach to characterize the sentiment. VADER uses gold-standard quality like Linguistic Inquiry and Word Count (LIWC) [22], which has been validated by humans. It distinguishes itself from other efficient tools, such as LIWC, in that it is more sensitive to sentiment expression in social media contexts. VADER also provides valence of the sentiment on a range from 1 to 9. Owing to the sentiment score, we can know through VADER the extent to which the sentiment is negative or positive.

This valence is based on generalizable rules that represent grammatical and syntactical conventions that humans use in contexts meant for emphasizing sentiment intensity. For our purpose, another important feature of VADER is the inclusion of sentiment-bearing lexical nonverbal items, such as emoticons, and verbal items, such as slang, acronyms, and initialisms, which are prevalent in social media contexts. The combination of valence polarity through both the lexicon and rule-based approaches is valuable for fine-grained sentiment analysis. The shortcomings of the lexicon-based approach include coverage, general sentiment intensity, and acquiring a new set of human lexical features.

Manual Coding and Analysis of Tweets

We stored tweets with negative sentiments separately to carry out further analysis. We assessed the themes and categories that most prominently featured in the tweets with negative sentiment. This we performed through manual coding and analysis. We composed a list where the number of occurrences of each word was presented. Before creating the list, we removed stop words and carried out lemmatization to find out the root words. Lemmatization reduces the number of words, and hence, the list was compact.

Manual analysis of the list of occurrences helped in devising larger socioeconomic or emotional-personal categories. The lists were thoroughly searched to identify meaningful words. There were words used for grammatical construction and others not mentioned significantly. This step was subjective and depended on the number of occurrences in the list. The words to include as topics under different categories were guided by the literature [21,23]. The creation of categories and the assignment of topics were subjective. This part of the classification can be considered qualitative as it was guided by the authors' judgments and opinions based on the literature.

Reddit

The Reddit assessment involved only 2 steps: data collection and manual coding. The intermediary step was not required because the Reddit data were forum-based, with a focus on coping with the negative effects of loneliness.

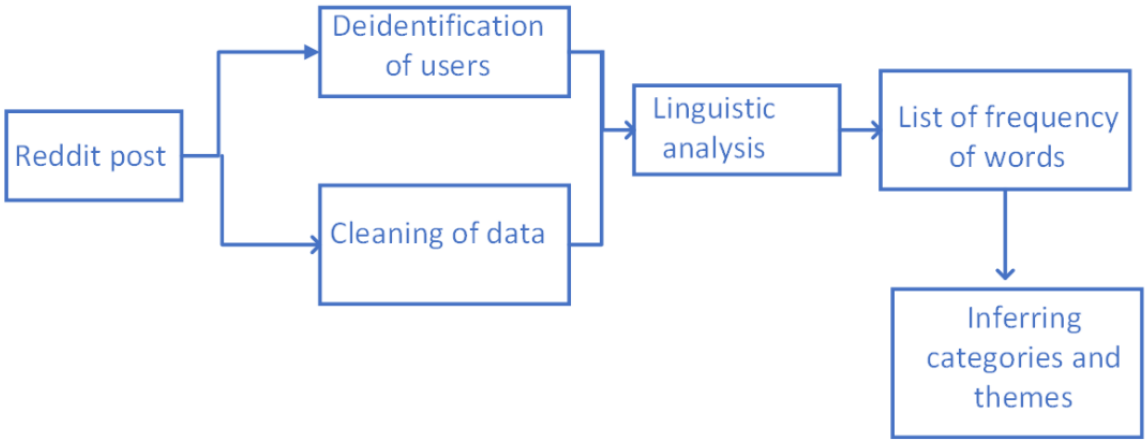
Reddit Data Collection

The Reddit data collection methodology was relatively straightforward. Reddit is a forum-based social media platform where people post about a topic on a subforum dedicated to the topic. These subforums are called subreddits. The Reddit API provides access to individual subreddits to download posts along with comments. The API provides access to download the top posts on a topic determined by the number of up-votes and other parameters of engagement. Figure 3 presents our pipeline for processing posts from Reddit. The difference between Figure 1 and Figure 3 is that there is no additional step of sentiment analysis for Reddit posts. After glossing over the subreddit "loneliness," we found that the posts were about the emotional expression of loneliness and did not involve metaphorical or off-topic use of loneliness. The nature of Reddit and subreddits designed for each topic encourages serious engagement on the topics. Because of this fact, sentiment analysis of Reddit data was not deemed important, and posts were analyzed through the frequency of the occurrence of words to identify the themes and topics most highly associated with loneliness.

Reddit posts from the "loneliness" subreddit were collected through the Reddit API. The "loneliness" subreddit had 13,000 members, and people could post and comment in the subforum. Reddit has its own algorithm for providing a score (ie, higher visibility for posts), which also considers inputs from other users in the form of up-voting. We collected the top 2000 Reddit posts from the "loneliness" subforum with all comments. The comments varied for each post in terms of both number and size. It is worth noting that some of the comments were of the same or even larger size than the original post. Thus, comments can be considered to include data on loneliness. The total number of individual texts analyzed was more than 2000 (posts multiplied by the average number of comments). Some posts did not have any comments, and the maximum number of comments for a post in the data we collected was 55. The average number of comments was 4.51, while the total number of comments was 8570. Combining the comments with the posts provided more than 10,000 unique texts or personal expressions on loneliness from Reddit.

XSL•FO
RenderX

Figure 3. Pipeline for processing Reddit data.



Reddit Data Manual Coding and Analysis

We analyzed both the posts and comments to identify the frequency of the occurrence of words to locate the associations of topics and themes with loneliness. The Reddit posts and comments were analyzed without performing any sentiment analysis.

The Reddit posts and comments were collected in a database, and stop words were removed. We performed lemmatization to find out the root words. Then, a list of occurrences of each word was created and analyzed similarly to the method presented in the subsection Manual Coding and Analysis of Tweets.

Reason for Using Manual Coding and Analysis

As the aim of this study was to identify differences between expressions of loneliness on Twitter and Reddit, assessment of important topics was an objective way to achieve this. A qualitative comparison of different tweets and different Reddit posts could be performed, but this would be biased and limited. When the comparison involves thousands of tweets and Reddit posts, it might not be possible to achieve a meaningful subjective comparison. Breaking down the tweets and Reddit posts into topics and categories allows for objective assessments of similarities and differences.

This method of searching for relevant categories of sociopolitical and personal-emotional content and topics was used because it has more flexibility. Usually, the n-gram method indicates the association of words that co-occur, which cannot be perceived to have happened by chance. However, this method does not indicate the occurrence of individual topics and words, and thus, the impact of a topic will not be known. The use of a keyword-based method (ie, using the collected data for some keywords and then reporting their occurrences) is a deductive approach, and it can help in searching the data for predefined keywords. In this study, we used more of an inductive approach, where we let all the important themes and topics in the data emerge. This allowed us to perform a more thorough analysis of data rather than being limited by keywords to look for in the data. This method to analyze the data is a quantitative approach. We did not qualitatively analyze the tweets and Reddit posts (ie, we did not provide our observations on what is contained in them). We reported the number of occurrences of meaningful words and then divided them into relevant categories. This quantitative stating of sentiment analysis and the words contained in negative sentiment tweets under categories was achieved by reading the literature and following the literature.

Results

Comparison of Reddit and Twitter Data

The number of tweets and Reddit posts for analysis was not determined by a specific rule. In the literature, different numbers of tweets and Reddit posts have been used for analysis. For example, Zubaiga [24] carried out an analysis of collected tweets to show the diversity of subjects and dataset sizes. The size of the datasets ranged from over a hundred thousand to more than 10 million. The aim of the study was to reflect in indicative terms the difference in the use of social media according to the expression of loneliness, and therefore, the numbers identified were considered reasonable.

Tables 1 and 2 provide basic details of Twitter and Reddit data, respectively. As discussed earlier, the dataset for Twitter analysis was obtained by combining tweets from the United States and India. The proportions of tweets with negative sentiment for the dataset of each country and for the combined dataset are presented in Table 1. Tweets with negative sentiment were then further analyzed to identify the frequency of the occurrence of topics. For Reddit data, the breakup of the data into words provided a total of more than 25,000 words. For the sake of the meaningful mention of topics and brevity, we only included topics that were mentioned at least 50 times, resulting in 411 words for analysis. Among these 411 words, there were many words related to language construction, such as propositions, and sentence structure. Words that were meaningful in terms of emotions or other expressive qualities were finally included in the analysis.

Table 3 presents the topics highly associated with negative mentions of loneliness in the Twitter dataset. Tweets with negative sentiment were first tokenized and stemmed to obtain a concise list of words and topics associated with loneliness. The list was then analyzed, and meaningful words representing topics of interest, such as emotional, social, and health identifiers, were found. Words like “oh,” “yeah,” and “ur” were ignored when composing the list. For the overall dataset, intimate relationships and interpersonal relationships showed

XSL•FO
RenderX

the highest associations with loneliness (Table 3). “Death” as a topic representing matters of health occurred the most in mentions related to loneliness, along with mentions related to COVID. In terms of socioeconomic factors and political expressions, a wide range of topics were identified.

Table 4 provides a list of themes mentioned with loneliness in the loneliness subreddit. The numbers given represent the total number of occurrences of the word or topic in the Reddit posts. The focus was mostly on relations and emotional expression. As longer posts are allowed, people have more space to express their feelings and open up. Social media platforms provide spaces to own one’s vulnerabilities without facing the backlash that can come in the form of social ostracization. In Table 4, the topics and their relevant thematic categories provide deeper insights into the personal-emotional associations of loneliness.

Table 1. Analysis of Twitter data.

Country	Sentiment analysis of tweets (negative sentiment proportion)
India	33.8%
United States	48.3%

Table 2. Analysis of Reddit data.

Variable	Value, n
Total words	35,057
Words occurring more than 100 times	611
Words occurring more than 1000 times	78

XSL•FO
RenderX

Table 3. Topics highly associated with mentions of loneliness in Twitter data.

Thematic area and topic	Mentions, n
Intimate relationships	
Family	2550
Children	1539
Cheat	64
Woman	3534
Relationship	1213
Interpersonal relationships	
Want	10,452
Need	9854
Feel	5173
Health	
COVID	5715
Die	9652
Life	3729
Socioeconomic factors	
Injustice	1538
Money	1344
Poor	3391
Culture	4452
Colorism (black/white)	2276
Emotional expression/insecurities	
Sad	2343
F*ck	5621
Hate	3542
Political	
War	1328
Protect	1198
Visa	2731
Modi	1576
Citizen	543
Custody	3742
Protest	4068
Insomnia	
Night	7154
Day	4315
Sleep	1564
Awake	543

XSL•FO
RenderX

Table 4. List of themes associated with the expression of loneliness in Reddit data.

Thematic area and topic	Value, n
Intimate relationships	
Love	563
Women	196
Relationship	233
Family	238
Single	111
Friends	1015
Social relations	
Friends	172
Girl	110
She	587
Her	467
People	146
Online	106
Meet	237
Person	428
Interpersonal relations	
Me	2648
He	101
Yours	1596
Us	196
Everyone	245
People	1775
Others	194
Emotional expression	
Thought	197
Hurt	110
Trying	284
Pain	114
Experience	105
Remember	101
Understand	268
Feeling	435
Want	941
Need	539
Feel	1904
Wish	229
Care	262
Self-focused	
I	13,604
Mental	119
My	3989

XSL•FO
RenderX

Thematic area and topic	Value, n
You	5424
Work-related	
Work	331
Job	117
Tried	190
Time	105
School	175
About time	
Life	1608
Year	234
Live	269
Old	145

Similarities and Differences in Reddit and Twitter Data

There are similarities in the categories and themes that can be found in conjunction with mentions of loneliness, but differences in the use of both social media platforms can be crucial in devising strategies to use social media data to understand loneliness. Reddit data are more focused on themes around personal-emotional categories and have a larger number of occurrences of topics around these categories. For Twitter, a range of categories were identified, and personal-emotional categories were the dominant ones, but other topics also occurred. Moreover, the associations of themes and topics with expressions of loneliness were in line with broader psychological linguistic features focused on the self-expression of mental health issues. The difference again was in the range. Twitter can have a wider range of topics that are associated with expressions of loneliness, while Reddit is more focused on psychological personal-emotional categories. These findings will help in providing a guideline for further research into deploying social media data to understand loneliness.

The differences that can be observed in Tables 3 and 4 involve diversity and extensiveness. In Twitter data, the range or diversity of topics and themes can be seen. Because of the limited-character expression on Twitter, people can express their thoughts or opinions, and from these, we can find the association with loneliness. There can be a range of such themes involving direct mentions of loneliness with mentions of a topic in a negative way. On the other hand, Reddit data indicate the extensiveness or depth of a theme or group of topics associated with loneliness.

Some categories, such as “political” and “insomnia,” are missing from Reddit data. The reason for this might be a conjuncture, as direct evidence for the lack of data on Reddit needs to be obtained after following a proper study protocol. However, it can be assumed that because Reddit is more personal (ie, anonymous with the culture being encouraged to reveal personal problems), the general associations of loneliness, such as political and extensive social themes, are not present. For “insomnia,” the easy nature of Twitter (ie, the requirement of short posts) might be seen as a better alternative to discuss sleep problems later at night. Again, these can be conjectural inferences, but the wider social and cultural context can inform the interpretation of these results.

Another reason for Twitter data containing socioeconomic and political categories, which are lacking in Reddit data, is the geographical composition of the data. However, further investigation is needed to determine the geographical composition of the data. According to these findings, Reddit data could be useful for finding the depth of a theme associated with loneliness (ie, what subthemes or topics under a broader category are related to loneliness). Twitter, on the other hand, showed the range and diversity of themes and categories associated with loneliness. This is important to investigate the possible causes of loneliness. The fact that Twitter provides a range of topics and themes and Reddit shows the depth of a theme can be used in a complimentary way.

Discussion

Principal Findings

The study performed a comparative analysis of 100,000 tweets and around 10,000 Reddit posts to understand the themes associated with the expression of loneliness according to socioeconomic and personal-emotional topics. The number of Reddit posts was kept lower than the number of tweets because Reddit posts are lengthier. These 2 social media platforms are important because of their unique nature. Twitter allows short texts, while Reddit allows a freehand composition of one’s opinion and analysis. Analysis of the tweets and Reddit posts provided attributes to understand the associations of socioeconomic and personal-emotional factors with loneliness. These attributes included emotion, sentiment, emojis, and topics. The analysis demonstrated that such attributes could help gather evidence and analyze interactions on the topic of loneliness and other such related topics. The first attribute was emotion, which can serve as a guide in understanding people’s reactions. The second most common attribute was relationships. Other thematic areas, such as health, work, self-focused topics, and

XSL•FO
RenderX

insomnia-related topics, indicated the intimate nature of loneliness. The major finding of this study was that both social media platforms differed in the range of expressions of themes in association with loneliness.

There were similarities in the themes associated with loneliness on social media, but differences in platform usage are crucial for strategies to understand loneliness. Reddit was heavily focused on personal-emotional themes, with a higher frequency of these topics. Twitter, while also emphasizing personal-emotional themes, included a broader range of categories. The thematic associations with loneliness aligned with broader psychological linguistic features centered on mental health self-expression. The key difference was range. Twitter had a wider variety of associated topics, whereas Reddit concentrated on personal-emotional themes. These findings provide guidelines for future research using social media data to understand loneliness.

The methodology developed in this paper showed the association of loneliness with language, which was associated with mental health issues such as anger and depression. The tweets and Reddit posts analyzed prove that psychosocial linguistic features can be found in the self-expression of loneliness, which can identify the dynamics of loneliness [25,26]. Tweets containing keywords associated with loneliness represent a self-focused discourse, which affirms previous literature on loneliness [27,28].

The topics and categories involve linguistic expressions of underlying mental health and related emotions. The findings of this paper regarding the topics and themes associated with loneliness, including mental health, alcoholism, emotional dysregulation, and sociopolitical circumstances, are backed by the literature. The relationships between mental health issues, such as depression and suicidal ideation, and issues related to neurobiology have been discussed by Lam et al [29]. These include conformity with the literature on the association of loneliness with emotional dysregulation and trouble with relationships [30]. The LIWC was used for different thematic concerns for loneliness by previous authors [23]. The study presented in tabular form the related social and emotional supercategories and relevant subcategories of loneliness. Self-referential pronouns were associated with loneliness according to the study. This study also affirms this finding. Our data present such linguistic categories, which have been proven by previous research to be associated with the linguistic nature of loneliness [31].

Finding themes associated with mental health and other issues in public health, such as loneliness, through social media analysis is interesting because such an analysis can provide insights into the first-hand experiences of people. The primary question in this study was “Is there a significant difference in the use of different platforms when it comes to the expression of loneliness?” The answer to this question is not binary. When it comes to larger patterns of the association of topics with loneliness, both platforms can provide similar patterns. However, the difference lies in the range of topics expressed and the depth of a particular topic. This finding is in line with previous findings. A previous study found that Twitter and Reddit can provide different sentiment expressions after a while for opinions regarding vaccination [32].

Which platform is better to use for the analysis of topics associated with loneliness and other related public health issues? From the data analysis and results, we found that Twitter can provide access to a large number of tweets through its API, but special steps in data cleaning and the combination of different data frames have to be taken in order to carry out a meaningful analysis. Reddit, on the other hand, provides access to lengthier posts where users are free from the constraints of space, and multiple insights into the nature of loneliness can be gained from a single post. Therefore, Twitter provides access to a range of insights, while Reddit provides access to a depth of insights. We also found that Reddit posts, although lesser than tweets, provided a lot more topics, but the topics were constrained by the variety of themes and belonged to similar thematic areas.

This study has provided the following insights into the nature and dynamics of loneliness through the collection of data from different social media platforms (Twitter and Reddit):

1. Despite the limitations of the dataset, the varying nature of loneliness can be observed on both platforms. It gives investigators the opportunity to approach the data on the expression of loneliness on different social media platforms differently.
2. Reddit allows users to share their opinions without limits, and Reddit is built around communities. These communities allow people having similar experiences to share their stories in detail. Therefore, Reddit is more suitable for educational and other messaging services, while Twitter can provide a range of topics, which can further be investigated in depth.
3. The data returned by Twitter can be very large, and to use the data, special data cleaning tools and other steps, such as sentiment analysis, should be used. Reddit data are from communities, and as such, they can be directly analyzed. There can be posts from bots and other such automatic or malicious agents. We did not remove these in this study. These need to be removed before further analysis in future work.

Limitations

This study has a number of limitations. First, the data from Twitter and Reddit are self-reported and may not represent the broader population. Only English posts were analyzed, limiting generalizability. Future research should explore various platforms and languages for a more comprehensive understanding of loneliness. Second, the dataset size of both Twitter data and Reddit data was small. We used a limited dataset to determine whether loneliness has different expressions on these platforms. The point was not to be exhaustive but indicative. This limitation can be addressed in a future study by increasing the dataset size to verify or change the findings of this study. Third, the Reddit data were not geo-tagged, and thus, we were not able to determine the countries of the users. However, there are methods to identify the locations of the users of Reddit posts. In the future, these methods can be used to carry out a comparative analysis with geo-located Twitter data. Fourth, there was no control over the data returned by the

XSL•FO
RenderX

Twitter API regarding specifying a country for data collection. The data returned was not probabilistically sampled, and thus, there may be bias in terms of both content and the country or region from which the data are collected. Fifth, the sentiment analysis of Twitter data may have issues. The automatic classification of tweets into negative and positive through sentiment analysis may not be fully accurate. While this was the basis of the study to carry out automated analysis, the results of this automated sentiment analysis need to be validated by looking at a certain number of tweets that were identified as negative. Current models do not detect some linguistic expressions such as sarcasm and humor. New models can be developed, which can correctly classify the sentiment of tweets that use sarcasm and humor. With this approach, we will be able to know the confidence of the analysis and quantify errors. This is related to the methodology of the study, which did not analyze tweets with negative or neutral sentiments. There is a possibility that even a positive or neutral mention of loneliness is associated negatively with mental health, but for the sake of a coherent analysis of a large number of tweets, this has been removed. Validation of sentiment analysis is needed to determine how many tweets with positive sentiment actually had negative connotations.

Conclusion

This study aimed to compare data on loneliness from Twitter and Reddit to identify topics and themes associated with the expression of loneliness. The findings of this study suggest that Twitter provides a wide and diverse range of data related to loneliness, while Reddit shows the depth or extensiveness of a theme associated with loneliness. By analyzing the language used in tweets and Reddit posts, it was possible to identify common themes associated with loneliness, such as social isolation, mental health, and relationship issues. The results of this study provide valuable insights into the ways in which people express and experience loneliness on social media platforms. The ability to identify these themes and topics associated with loneliness could be useful for health professionals and researchers to understand the complex nature of loneliness and its impact on individuals.

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

HS designed the study, collected the data, carried out data analysis, and wrote the paper. MH designed and supervised the study.

Conflicts of Interest

None declared.

References

1. Shah S, Nogueras D, van Woerden H, Kiparoglou V. Evaluation of the effectiveness of digital technology interventions to reduce loneliness in older adults: Systematic review and meta-analysis. J Med Internet Res 2021 Jun 04;23(6):e24712 [FREE Full text] [doi: 10.2196/24712] [Medline: 34085942]
2. Hards E, Loades ME, Higson-Sweeney N, Shafran R, Serafimova T, Brigden A, et al. Loneliness and mental health in children and adolescents with pre-existing mental health problems: A rapid systematic review. Br J Clin Psychol 2022 Jun 16;61(2):313-334. [doi: 10.1111/bjc.12331] [Medline: 34529837]
3. Peplau L, Perlman D. Loneliness: A Sourcebook of Current Theory, Research and Therapy. New York, NY: Wiley; 1982.
4. Achterbergh L, Pitman A, Birken M, Pearce E, Sno H, Johnson S. The experience of loneliness among young people with depression: a qualitative meta-synthesis of the literature. BMC Psychiatry 2020 Aug 24;20(1):415 [FREE Full text] [doi: 10.1186/s12888-020-02818-3] [Medline: 32831064]
5. Shah HA, Househ M. Understanding loneliness in younger people: Review of the opportunities and challenges for loneliness interventions. Interact J Med Res 2023 Nov 02;12:e45197 [FREE Full text] [doi: 10.2196/45197] [Medline: 37917125]
6. Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. In: Proceedings of the International AAAI Conference on Web and Social Media. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, Michigan, USA p. 216-225. [doi: 10.1609/icwsm.v8i1.14550]
7. Wiwatkunupakarn N, Pateekhum C, Aramrat C, Jirapornchaoren W, Pinyopornpanish K, Angkurawaranon C. Social networking site usage: A systematic review of its relationship with social isolation, loneliness, and depression among older adults. Aging Ment Health 2022 Jul 24;26(7):1318-1326. [doi: 10.1080/13607863.2021.1966745] [Medline: 34427132]
8. Choi HK, Lee SH. Trends and effectiveness of ICT interventions for the elderly to reduce loneliness: A systematic review. Healthcare (Basel) 2021 Mar 07;9(3):293 [FREE Full text] [doi: 10.3390/healthcare9030293] [Medline: 33800099]
9. Khurana A, Rastogi R, Gamperl HJ. A new era of drug safety - New EU pharmacovigilance (PV) legislation and comparison of PV in EU, US and India. International Journal of Pharmacy and Pharmaceutical Sciences 2014;6(7):15-21 [FREE Full text]
10. Sathyanarayanan V, Flowers CR, Iyer SP. Comparison of access to novel drugs for lymphoma and chronic lymphocytic leukemia between India and the United States. JCO Global Oncology 2020 Nov(6):1124-1133. [doi: 10.1200/go.20.00012]

XSL•FO
RenderX

11. Foufi V, Timakum T, Gaudet-Blavignac C, Lovis C, Song M. Mining of textual health information from Reddit: Analysis of chronic diseases with extracted entities and their relations. J Med Internet Res 2019 Jun 13;21(6):e12876 [FREE Full text] [doi: 10.2196/12876] [Medline: 31199327]
12. Schrading N, Alm C, Ptucha R, Homan C. An Analysis of Domestic Abuse Discourse on Reddit. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. 2015 Presented at: 2015 Conference on Empirical Methods in Natural Language Processing; September 17-21, 2015; Lisbon, Portugal p. 2577-2583. [doi: 10.18653/v1/d15-1309]
13. Tsai M, Wang Y. Analyzing twitter data to evaluate people's attitudes towards public health policies and events in the era of COVID-19. Int J Environ Res Public Health 2021 Jun 10;18(12):6272 [FREE Full text] [doi: 10.3390/ijerph18126272] [Medline: 34200576]
14. Valdez D, Ten Thij M, Bathina K, Rutter LA, Bollen J. Social media insights into US mental health during the COVID-19 pandemic: Longitudinal analysis of twitter data. J Med Internet Res 2020 Dec 14;22(12):e21418 [FREE Full text] [doi: 10.2196/21418] [Medline: 33284783]
15. Ahmed W, Vidal-Alaball J, Vilaseca Llobet JM. Analyzing discussions around rural health on twitter during the COVID-19 pandemic: Social network analysis of twitter data. JMIR Infodemiology 2023 Mar 8;3:e39209 [FREE Full text] [doi: 10.2196/39209] [Medline: 36936067]
16. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: Machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
17. Cinelli M, De Francisci Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2023301118 [FREE Full text] [doi: 10.1073/pnas.2023301118] [Medline: 33622786]
18. Curiskis SA, Drake B, Osborn TR, Kennedy PJ. An evaluation of document clustering and topic modelling in two online social networks: Twitter and Reddit. Information Processing & Management 2020 Mar;57(2):102034. [doi: 10.1016/j.ipm.2019.04.002]
19. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective response to media coverage of the COVID-19 pandemic on Reddit and Wikipedia: Mixed-methods analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
20. Babvey P, Capela F, Cappa C, Lipizzi C, Petrowski N, Ramirez-Marquez J. Using social media data for assessing children's exposure to violence during the COVID-19 pandemic. Child Abuse Negl 2021 Jun;116(Pt 2):104747 [FREE Full text] [doi: 10.1016/j.chiabu.2020.104747] [Medline: 33358281]
21. Guntuku SC, Schneider R, Pelullo A, Young J, Wong V, Ungar L, et al. Studying expressions of loneliness in individuals using twitter: an observational study. BMJ Open 2019 Nov 04;9(11):e030355 [FREE Full text] [doi: 10.1136/bmjopen-2019-030355] [Medline: 31685502]
22. Pennebaker J, Francis M, Booth R. Linguistic Inquiry and Word Count. Mahwah, NJ: Erlbaum Publishers; 2001.
23. Liu T, Ungar LH, Curtis B, Sherman G, Yadeta K, Tay L, et al. Head versus heart: social media reveals differential language of loneliness from depression. Npj Ment Health Res 2022 Oct 18;1(1):16 [FREE Full text] [doi: 10.1038/s44184-022-00014-7] [Medline: 38609477]
24. Zubiaga A. A longitudinal assessment of the persistence of twitter datasets. Asso for Info Science & Tech 2018 May 14;69(8):974-984. [doi: 10.1002/asi.24026]
25. Stravynski A, Boyer R. Loneliness in relation to suicide ideation and parasuicide: a population-wide study. Suicide Life Threat Behav 2001 Jan 07;31(1):32-40. [doi: 10.1521/suli.31.1.32.21312] [Medline: 11326767]
26. Blai B. Health consequences of loneliness: a review of the literature. J Am Coll Health 1989 Jan 09;37(4):162-167. [doi: 10.1080/07448481.1989.9938410] [Medline: 2647810]
27. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct 22;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
28. Edwards T, Holtzman NS. A meta-analysis of correlations between depression and first person singular pronoun use. Journal of Research in Personality 2017 Jun;68:63-68. [doi: 10.1016/j.jrp.2017.02.005]
29. Lam JA, Murray ER, Yu KE, Ramsey M, Nguyen TT, Mishra J, et al. Neurobiology of loneliness: a systematic review. Neuropsychopharmacology 2021 Oct 06;46(11):1873-1887 [FREE Full text] [doi: 10.1038/s41386-021-01058-7] [Medline: 34230607]
30. Al-Saggaf Y, Nielsen S. Self-disclosure on Facebook among female users and its relationship to feelings of loneliness. Computers in Human Behavior 2014 Jul;36:460-468. [doi: 10.1016/j.chb.2014.04.014]
31. Buecker S, Ebert T, Götz FM, Entringer TM, Luhmann M. In a lonely place: Investigating regional differences in loneliness. Social Psychological and Personality Science 2020 Apr 10;12(2):147-155. [doi: 10.1177/1948550620912881]
32. Melton CA, White BM, Davis RL, Bednarczyk RA, Shaban-Nejad A. Fine-tuned sentiment analysis of COVID-19 vaccine-related social media data: Comparative study. J Med Internet Res 2022 Oct 17;24(10):e40408 [FREE Full text] [doi: 10.2196/40408] [Medline: 36174192]

XSL•FO
RenderX

Abbreviations

API: application programming interface
LIWC: Linguistic Inquiry and Word Count
NLTK: natural language toolkit
SIA: social intelligence analysis
VADER: valence aware dictionary for sentiment reasoning

Edited by T de Azevedo Cardoso; submitted 30.05.23; peer-reviewed by C Entwistle, T Cahill; comments to author 28.08.23; revised version received 29.10.23; accepted 25.07.24; published 14.03.25.

Please cite as:
Shah HA, Househ M
Understanding Loneliness Through Analysis of Twitter and Reddit Data: Comparative Study
Interact J Med Res 2025;14:e49464
URL: https://www.i-jmr.org/2025/1/e49464
doi:10.2196/49464
PMID:

©Hurmat Ali Shah, Mowafa Househ. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.03.2025. 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

Exploring 97 Years of Aedes aegypti as the Vector for Dengue, Yellow Fever, Zika, and Chikungunya (Diptera: Culicidae): Scientometric Analysis

Yoon Ling Cheong[1], PhD; Sumarni Mohd Ghazali[2], MSc; Mohd Hazilas Mat Hashim[1], Dip FHS; Mohd Khairuddin Che Ibrahim[3], BSc; Afzufira Amran[1], BSc; Tsye Yih Tiunh[1], MSc; Hui Li Lim[4], MSc; Yong Kang Cheah[5], PhD; Balvinder Singh Gill[6], PhD; Kuang Hock Lim[1], MSc

[1]Biomedical Museum Unit, Special Resource Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, Kuala Lumpur, Malaysia

[2]Biomedical Epidemiology Unit, Special Resource Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, Selangor, Malaysia

[3]Biomedical Research, Strategic & Innovation Management Unit, Institute for Medical Research, Director's Office, National Institutes of Health, Ministry of Health Malaysia, Selangor, Malaysia

[4]Clinical Research Centre, National Institutes of Health, Ministry of Health Malaysia, Selangor, Malaysia

[5]School of Economics, Finance & Banking, Universiti Utara Malaysia, Kedah, Malaysia

[6]Special Resource Centre, Institute for Medical Research, National Institutes of Health, Ministry of Health Malaysia, Selangor, Malaysia

Corresponding Author:
Yoon Ling Cheong, PhD
Biomedical Museum Unit, Special Resource Centre
Institute for Medical Research, National Institutes of Health
Ministry of Health Malaysia
Jalan Pahang
Kuala Lumpur, 50588
Malaysia
Phone: 60 326162666 ext 2810
Email: cheongyl@moh.gov.my

Abstract

Background: *Aedes aegypti* is an important vector that transmits dengue, Zika, chikungunya, and yellow fever viruses. Although research on *Aedes aegypti* has been conducted for decades, scientometric studies on *Aedes aegypti* are scarce, are limited to regions, and cover short periods. Thus, there is still a knowledge gap in the current trend, research focuses and directions, leading authors and collaboration, journal and citation impacts, countries, and worldwide collaborations.

Objective: The objectives of the study are to investigate the research trend, focus and directions, citation impact, leading authors and collaboration, journals, and countries of the published works on *Aedes aegypti* to inform the current knowledge gaps and future direction of the control of the vector.

Methods: In this study, we searched the Scopus database for articles on *Aedes aegypti* published from the year 1927 until April 5th, 2024, and included articles, reviews, books, and book chapters that were written in English. A total of 16,247 articles in 160 journals with 481,479 citations were included. Inconsistencies in authors' names were checked and cleaned using OpenRefine. The data were grouped into 4 periods; years 1927-1999, 2000-2009, 2010-2019, and 2020-2023. The relative growth rate and doubling time of publications were calculated. The analysis was conducted using VOSviewer, R bibliometrics, and citeSpace.

Results: The overall RGR was 0.1. Doubling time increased from 9.3 in 1978-1998 to 12.1 in 2000-2009. The main research clusters were "using Wolbachia," "Dengue Zika," "worldwide diversity," "community support," "larvicidal activity," "mosquito genotype-dependent," and "sterile insect technique." *Journal of Medical Entomology* was the leading journal (758/16,247, 4.7%). The most cited articles were authored by Halstead SB and team in *Science* (N=1355) and Kraemer MU and team in *eLife* (N=1324). The United States (5806/23,538, 24.7%) and Brazil (2035/23,538, 8.6%) were the top countries. Gubler DJ was the top co-cited author (n=2892) from 2000 to 2019. The co-cited author cluster patterns informed the significant specialty research on *Aedes aegypti* across time. Authors from various specialized research fields tended to collaborate across countries, especially neighboring countries. Countries with more research funding on the study of *Aedes aegypti* published more papers.

XSL•FO
RenderX

Conclusions: Researchers or entomologists could understand the current knowledge gap on *Aedes aegypti* and plan for future research pathways. This study contributed to the public health stakeholders in improving the vector control interventions and elucidated the extent of research subject areas.

(Interact J Med Res 2025;14:e65844) doi:10.2196/65844

KEYWORDS

relative growth rate; vector control; collaboration index; bibliometric; Aedes aegypti; Scopus; co-occurrence; author; dengue; Zika

Introduction

Aedes aegypti is a vector of several arboviruses, including dengue, Zika, chikungunya, and yellow fever. This species was discovered by Fredrik Hasselqvist in Egypt in 1757 [1]. Initially designated as Culex aegypti, the name was declared invalid by the International Commission on Zoological Nomenclature in 1956 and it was subsequently identified as Stegomyia fasciata followed by *Aedes aegypti* (Linnaeus) or *Aedes* (Stegomyia) *aegypti* (Linnaeus) [1]. Its distinguishing features are the presence of white rings around the leg articulations and abdomen and a white lyre-shaped marking on the dorsal surface of the thorax [2]. Originally native to sub-Saharan Africa, *Aedes aegypti* is now widespread in most tropical and subtropical locations across the world [3]. The most common habitats of the *Aedes aegypti* are artificial and natural containers, including plastic containers, flower pots, vases, tires, tree holes, and bromeliad plants [4,5]. They are common in high-density residential areas [3]. *Aedes aegypti* can potentially survive and establish in 215 of 250 countries and territories (86%) [6].

Scientometrics concerns the analysis of subjects and the development of research literature, the impact and process of scientific knowledge production, as well as the macroenvironment of research [7]. The aim is to discover hidden relationships between any single publication and citations [8]. Studies on animal species using the scientometric method revealed trends in the direction of research on animal species. Santos and Vianna [9], who conducted a scientometric analysis of the literature on 11 Western Atlantic species of Paralichthys in ISI Web of Science and SciELO, discovered the dominant fishery species of *Paralichthys dentatus* (46.1%) in Canada and *Paralichthys lethostigma* (32.1%) in the United States. Araújo [10] discovered the average number of total citations per paper was 16.05 (27.66) in the scientific literature on insect galls and galling species in Brazil from 1988 to 2017. Miguel, Calvão, Vital, and Juen [11], who examined the literature on insects of the order Odonata, found the majority of the study focused primarily on the adult stage and species level, and there were gaps in the biogeography, parasitism, competition within and across species, evolutionary and phylogenetic links, and studies of the larval stages.

To date, there is a lack of scientometric analysis of *Aedes aegypti* at the global level. Vega-Almeida et al [12] conducted a scientometric analysis of 5039 published articles about *Aedes aegypti* from the year 2006 to 2015 (10 years) in Scopus. The findings were distributed across 4 major domains, namely epidemiology, gene expression and biological control, larvicidal and insecticidal effects, and reproduction and insecticide resistance; however, the paper did not analyze the information on authorship, citation rates, journal, organization, and collaborations. Other studies focused solely on the diseases transmitted by *Aedes aegypti* [13], or the effects of climate change on *Aedes aegypti* [14]. Gupta and Tiwari [15] discovered from the 910 articles publications of dengue research in India for the duration of the year 2003 to 2012, with an average annual growth rate of 28.19%. A scientometric analysis of the Zika virus, yielded 567 publications, with the most productive countries being the United States, the United Kingdom, and the Netherlands [16]. Sindhania et al [17] focused on the publication on *Aedes aegypti* and *Aedes albopictus* extracted from the Web of Science, encompassing a collection of 4149 papers for 77 years. Focus on single species was crucial as the targeted vector controls of both species were differed due to the uniqueness of their preferable habitats, morphology, genetic profile, and virus replication characteristics [18,19]. Hence, a scientometric study on *Aedes aegypti* based on Scopus that covers a wider aspect, including author collaborations and cocitations, citation rate and impact, leading journals, and countries would enable researchers and entomologists to map the current knowledge gap on *Aedes aegypti* and to plan future research focus or priorities. The Scopus offers more multidiscipline contents than Web of Science [20]. For public health practitioners, this information would assist in improving vector control interventions and evaluating the extent of past and potential research areas.

This study aims to conduct a scientometric analysis of scientific articles on *Aedes aegypti* published worldwide from 1927 to 2023. The specific objectives are numerous; first, to analyze the general trends and annual growth rate of the articles published on *Aedes aegypti*; second, to identify prolific authors, their collaborative networks, and authorship patterns; third, to determine the co-occurrences, patterns of subject areas, and current research trends; fourth, to study the journals and their citation impact; and finally, to determine the countries of the published papers.

Methods

Data Source and Search Strategy

We conducted our search in the Scopus database for publications from the year 1927 to 2023 using the search term “Aedes aegypti” on April 15, 2024 (Textbox 1). The following search queries were used in the investigation in Scopus in two phases. First, search all the publications up to the year 2023.

Second, search the publications by 4 time periods, that was years 1927-1999, 2000-2009, 2010-2019, and 2020-2023.

XSL•FO
RenderX

Textbox 1.

(TITLE-ABS-KEY("Aedes aegypti") AND PUBYEAR > 1926 AND PUBYEAR < 2024 AND (LIMIT-TO (DOCTYPE,"ar") OR LIMIT-TO (DOCTYPE,"re") OR LIMIT-TO (DOCTYPE,"ch") OR LIMIT-TO (DOCTYPE,"bk")) AND (LIMIT-TO (LANGUAGE,"English"))).

The search was conducted within the title, abstract, and keywords fields and was limited to articles, reviews, books, and book chapters in English. Documents that did not match the inclusion criteria were excluded. The excluded articles included retracted articles, conference papers, letters, notes, editorials, short surveys, erratum, conference reviews, and data reviews. Data extracted from the search results included publication year, title, author, citation, keywords, organization, journal, and country, and were saved in 5 "csv" format files. Inconsistencies in authors' names were checked and cleaned using OpenRefine [21]. For quality control, a random sample of 10% of the extracted data was verified against the original source publication. Data merging, pivoting, and aggregation analysis were conducted with Microsoft Excel.

Data Analysis

Relative Growth Rate and Doubling Time

We calculated the relative growth rate (RGR) and doubling time of publications. RGR, which measures the change in the number of publications per year, was calculated using the following formula:



where, t_1 is the initial time period, t_2 is the final time period, N_1 denotes the number of publications at time t_1, and N_2 denotes the number of publications at time t_2.

Given the RGR, we computed the doubling time. Doubling time is the time it takes for the number of publications to double.



Authorship Analysis

Analysis of authorship and subject area and construction of the respective network graphs were conducted using VOSviewer software developed by Leiden University [22]. The coauthorship analysis was based on the full counting method, that was, each coauthor of a publication is assigned a weight of one; thereby, the total weight of a publication is equal to the number of its authors. Author contribution timelines, 3-field plots, and country collaboration maps were generated using the R bibliometrics package [23].

Impact Factor and Cocitation Analysis

Journal publications' 5-year impact factor was extracted from Web of Science Journal Information. Cocitation analysis tracks pairs of study that are cited together and suggest similarity as both items were cited by the same study, which helps in identifying the inner structure of research disciplines that represent the research focus [24]. The analysis was conducted using the VOSviewer and citeSpace [25]. The author's cocitation analysis was split into 4 time periods: years 1927-1999, 2000-2009, 2010-2019, and 2020-2023. To understand the recent trend of research focus, the citeSpace analysis zoomed in on the year 2021-2023. The higher silhouette score of a cluster means members in the cluster have more in common in terms of tightness and separation [26].

The current study focused solely on *Aedes aegypti*, with *Aedes albopictus* being the next target of investigation. This approach is due to the species-specific nature of biological control and the differing habitats of both species. Conducting a scientometric analysis on a single species, provides a clearer picture, which can later be expanded to compare trends between the 2 species.

Ethical Considerations

Ethical approval protocol from the Medical Research and Ethics Committee, Ministry of Health Malaysia was exempted (reference number KKM/NIH/22-01333-RCQ[1]).

Results

General Trends

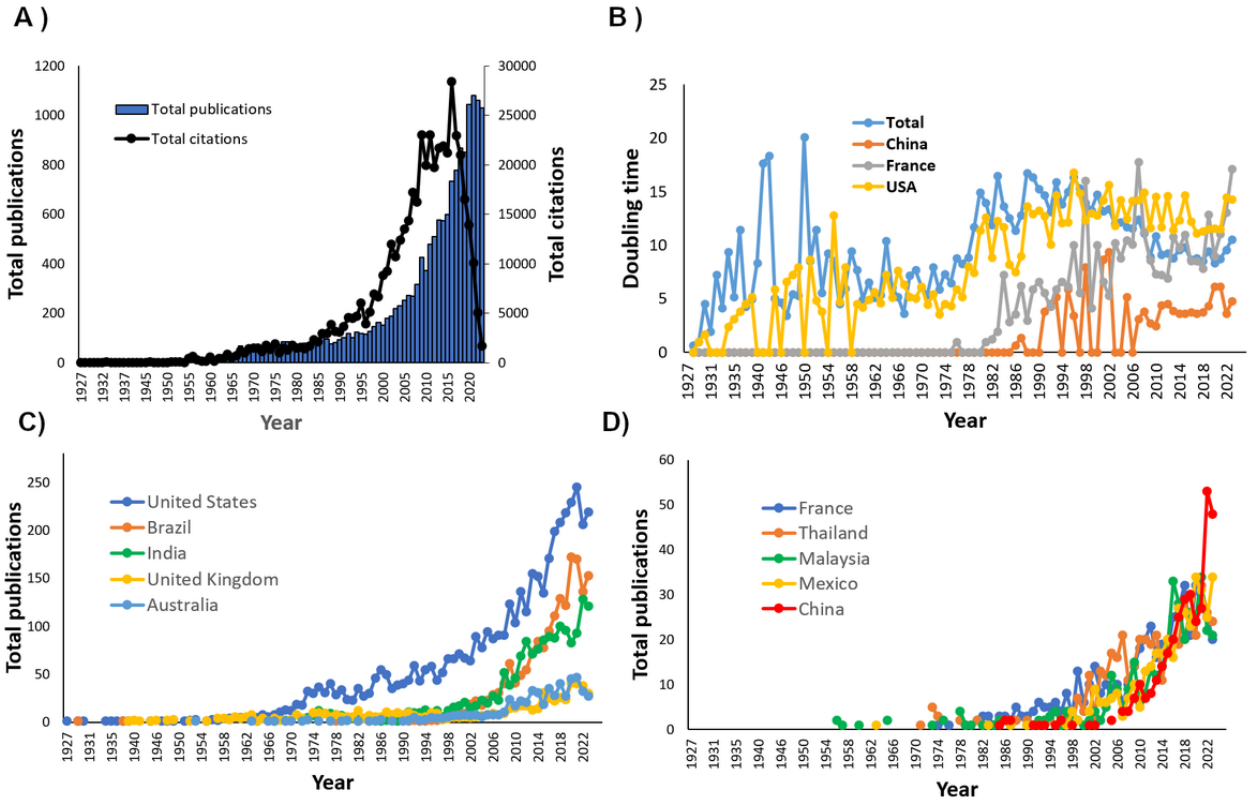
The search strategy produced a total of 18,529 results, mainly in English (17,438/18,686, 93.3%), Spanish (506/18,686, 2.7%), Portuguese (278/18,686, 1.5%), French (203/18,686, 1.1%), and the other languages. For the articles in English (n=17,438), the document types were research articles (14,990/17,438, 86.0%), followed by review articles (1063/17,438, 6.1%), conference proceedings (366/17,438, 2.1%), letters (253/17,438, 1.5%), notes (191/17,438, 1.1%), and others. Finally, only research articles, reviews, letters, editorials, books, and book chapters were included (n=16,247). Figure 1A shows that articles were published at a steeply increasing trend with a total number of 14,990 (14,990/16,247, 92.26%). The review articles increased steadily and exhibited a doubling in number after the year 2014, contributing to a total number of 1063 papers (1063/16,247, 6.54%). There were 189 (189/16,247, 1.16%) book chapters and 5 (5/16,247, 0.03%) books on *Aedes aegypti*. The total number of articles in each of the time periods was 3183, 2506, 6345, and 4214, respectively. The number of authors expanded rapidly from 4490, 6459, and 19,495 to 18,064, respectively. The top countries published scientific articles on "Aedes aegypti" were the United States (5806/23,538, 24.7%), followed by Brazil (2035/23,538, 8.6%), India (1839/23,538, 7.8%), United Kingdom (1243/23,538, 5.3%), France (898/23,538, 3.8%), and Australia (779/23,538, 3.3%; Figure 1C). The other countries with a significant number of articles were Thailand (678/23,538, 2.8%), Mexico (548/23,538, 2.3%), China (541/23,538, 2.3%), and Malaysia (455/23,538, 1.9%), respectively (Figure 1D).

The overall RGR was 0.1. Doubling time increased from 9.3 in 1978-1998 to 12.1 in 2000-2009, and reverted back to 9.3 in 2020-2023 (Figure 1B). The RGR and DT varied across countries. Among the top 10 countries in total publications, France exhibited the longest doubling time, while China had the shortest doubling time. The Unites States has published

XSL•FO
RenderX

since 1927 and had a short doubling time initially, but the numbers reduced in recent years (Figure 1B).

Figure 1. Distribution by year of (A) total number of publications and citations on Aedes aegypti globally, (B) total and selected countries' publication doubling time, (C) total number of publications by the top 5 countries, and (D) total number of publications by the 6th-10th ranked countries.



Institutional Contributions

Authors affiliated with the Fundacao Oswaldo Cruz produced 3.4% of the overall *Aedes aegypti* research output, followed by those in the University of Florida (2.3%), CNRS Centre National de la Recherche Scientifique (1.8%), Institut Pasteur, Paris (1.7%), Colorado State University (1.6%), Mahidol University (1.6%), Centers for Disease Control and Prevention (1.5%), University of Notre Dame (1.5%), Universidade de São Paulo (1.5%), and University of California, Davis (1.5%).

Funding Agencies

There were 159 funding agencies for the *Aedes aegypti* research whilst 8914/19,883 (50.2%) articles did not disclose their sources of funding. The National Institute of Allergy and Infectious Diseases, United States (1716/19,883; 9.7%) and National Institutes of Health, United States (1183/19,883; 6.67%) contributed to a quarter of the publications. Four agencies sponsored more than 200 articles each, that was, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (708/19,883, 4.0%); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil (576/19,883, 3.2%); National Science Foundation, United States (255/19,883, 1.4%); and Centers for Disease Control and Prevention, United States (255/19,883, 1.4%). For other regions and countries, the highest funders were National Natural Science Foundation of China (207/19,883, 1.2%); Wellcome Trust (158/19,883, 0.9%); National Health and Medical Research Council, Australia (149/19,883, 0.8%); Horizon 2020 Framework Programme, European Union (144/19,883, 0.8%); Department of Science and Technology, Ministry of Science and Technology, India (108/19,883, 0.6%); Natural Sciences and Engineering Research Council of Canada (78/19,883, 0.4%); Consejo Nacional de Ciencia y Tecnología, Mexico (77/19,883, 0.4%); and Agence Nationale de la Recherche, France (84/19,883, 0.5%).

Authorship Analysis

The top 20 productive authors who published articles related to *Aedes aegypti* originated from mixed countries, mainly the United States, for that were, Scott TW, Raikhel AS, Severson DW, Christensen BM, James AA, Becnel JJ, Tabanca N, Ali A, and Higgs S (Table 1). Three authors from Australia were among the first 10 productive authors, including Ritchie SA (n=126), Hoffmann AA (n=94), and O'Neill SL (n=79). The other authors were from France (Failloux A-B, n=95), Thailand (Chareonviriyaphap T, n=78), Italy (Benelli G, n=74), India (Govindarajan M, n=69), Malaysia (Lee HL, n=64), Mexico (Manrique-Saide P, n=63), Brazil (LourenÇo-de-oliveira R, n=55), and Trinidad and Tobago (Chadee DD, n=55). The author with the highest number of publications did not have the highest citation and H-index. The highest H-index author was Hoffmann AA (H-index=103) and the author with the highest total citation was Scott TW (n=14,100).

XSL•FO
RenderX

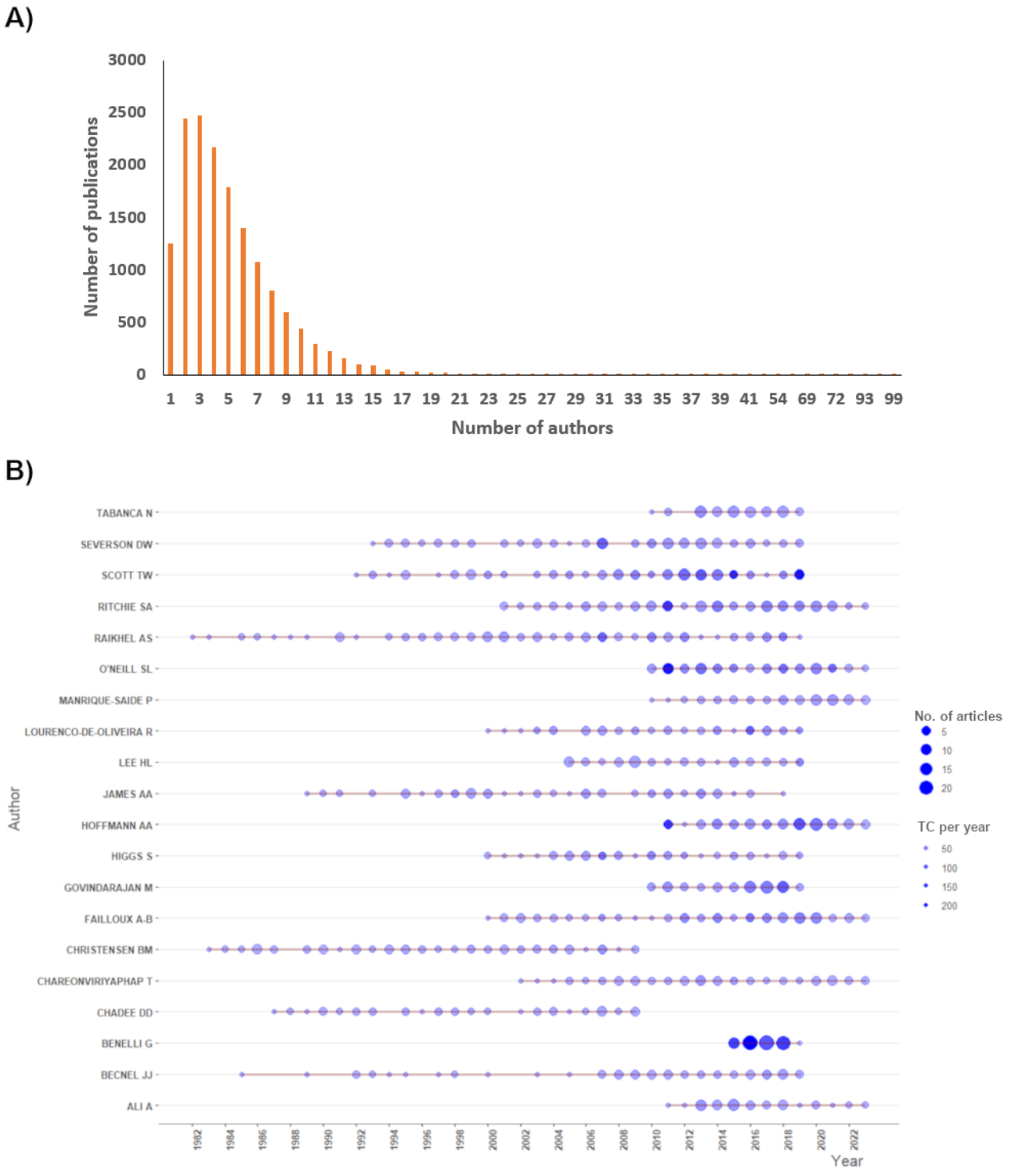
Table 1. Top 20 productive authors, citation, and Scopus H-Index.

Authors	Country	Total publication, n (%)	Total citation, n (%)	Scopus H-index (2024)
Ritchie SA	Australia	126 (0.78)	7522 (1.56)	59
Scott TW	United States	116 (0.71)	14,100 (2.93)	82
Raikhel AS	United States	115 (0.71)	9074 (1.88)	68
Severson DW	United States	108 (0.66)	5756 (1.20)	44
Failloux A-B	France	95 (0.58)	5370 (1.12)	53
Hoffmann AA	Australia	94 (0.58)	5659 (1.18)	103
Christensen BM	United States	88 (0.54)	3923 (0.81)	49
O'Neill SL	Australia	79 (0.49)	8050 (1.67)	77
Chareonnviriyaphap T	Thailand	78 (0.48)	1807 (0.38)	37
James AA	United States	77 (0.47)	5618 (1.17)	59
Benelli G	Italy	74 (0.46)	5394 (1.12)	84
Becnel JJ	United States	74 (0.46)	2026 (0.42)	43
Govindarajan M	India	69 (0.42)	3440 (0.71)	59
Tabanca N	United States	66 (0.41)	1545 (0.32)	38
Lee HL	Malaysia	64 (0.39)	1618 (0.34)	30
Manrique-Saide P	Mexico	63 (0.39)	1161 (0.24)	24
Ali A	United States	57 (0.35)	1232 (0.26)	22
Lourenço-de-Oliveira R	Brazil	55 (0.34)	4396 (0.91)	49
Chadee DD	Trinidad and Tobago	55 (0.34)	1855 (0.39)	35
Higgs S	United States	54 (0.33)	5149 (1.07)	67

The majority of the publications had between 2 and 5 authors. The number of publications decreased exponentially with increases in number of coauthors. There were more publications by multiple authors than by single authors (Figure 2A). The top-ranked authors were Raikhel AS, Christensen BM, Becnel JJ, and Chadee, each of whom has published articles on *Aedes aegypti* since 1982 (Figure 2B). Benelli G published several articles with high citations in a span of 4 years. Hoffmann AA, Manrique-saide P, Ritchie SA, O'Neill SL, Failloux AB, Chareonviriyaphap T, and Ali A are still actively publishing study on *Aedes aegypti* (up to 2023; Figure 2B).

XSL•FO
RenderX

Figure 2. Author analysis. (A) Number of publications by number of authors, and (b) top 20 authors' production over time and total citations per year.



In the early years, of all the 8975 cited authors, Raikhel AS had the highest number of cocitation links with other authors (n=303), followed by Hagedorn HH (n=248) and James AA (n=204; Figure 3A). From 2000 through 2009, of all the 71,005 cited authors, Gubler DJ was the most cocited author (n=1272), followed by Beaty BJ (n=711) and Raikhel AS (Figure 3B). From 2010-2019, Gubler DJ remained as the top co-cited author (n=2892) among 219,499 cited authors. Scott TW was the second co-cited author (n=2885), while Beneli G was the third (n=1986; Figure 3C). From 2020 to 2023, out of 230,547 cited authors, Scott TW (n=1311) and Hoffmann AA (n=858) stayed the highest in the cocitation list, with Brady OJ (n=1086) came in third (Figure 3D). The varied color depicted authors' linkages with different clusters. Gubler DJ, Brady OJ, and Scott TW were in the same cluster, whereas Hoffmann AA and Ritchie SA were in another cluster.

XSL•FO
RenderX

Figure 3. Cocitations network of authorship for the period of (A) 1927-1999, (B) 2000-2009, (C) 2010-2019, and (D) 2020-2023.

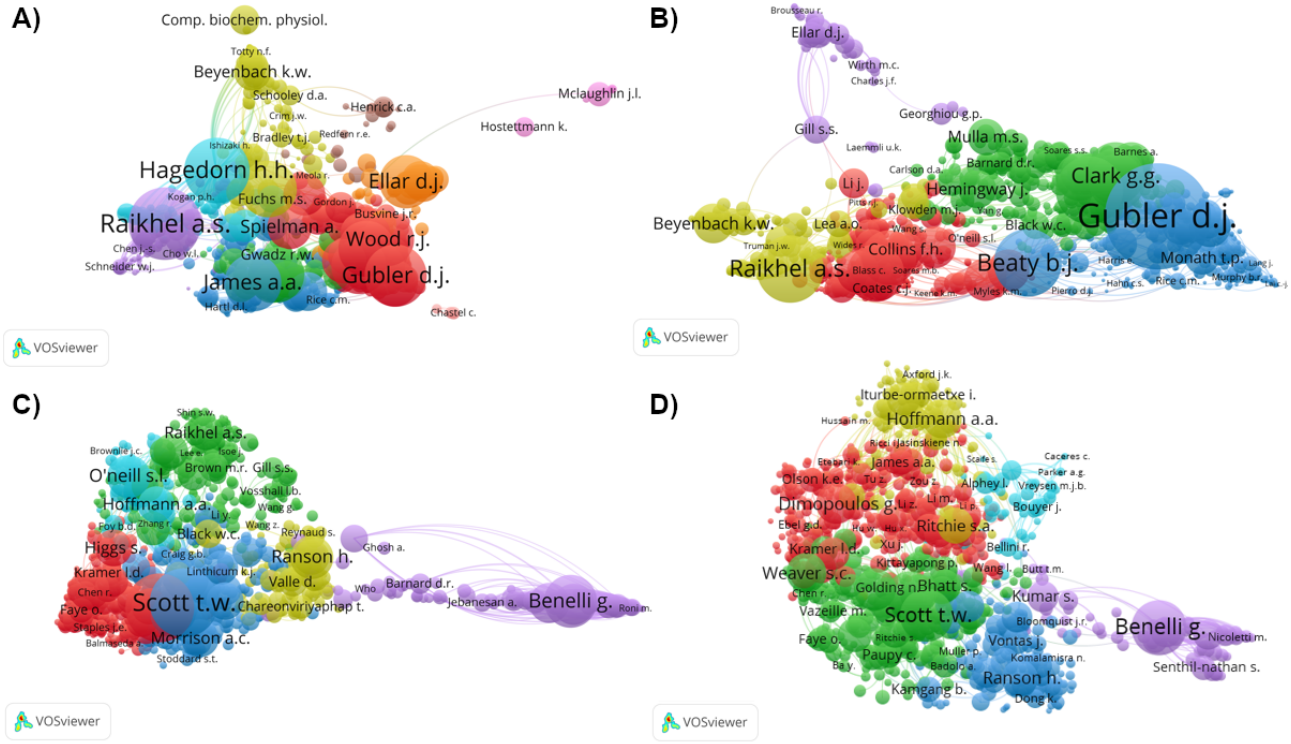
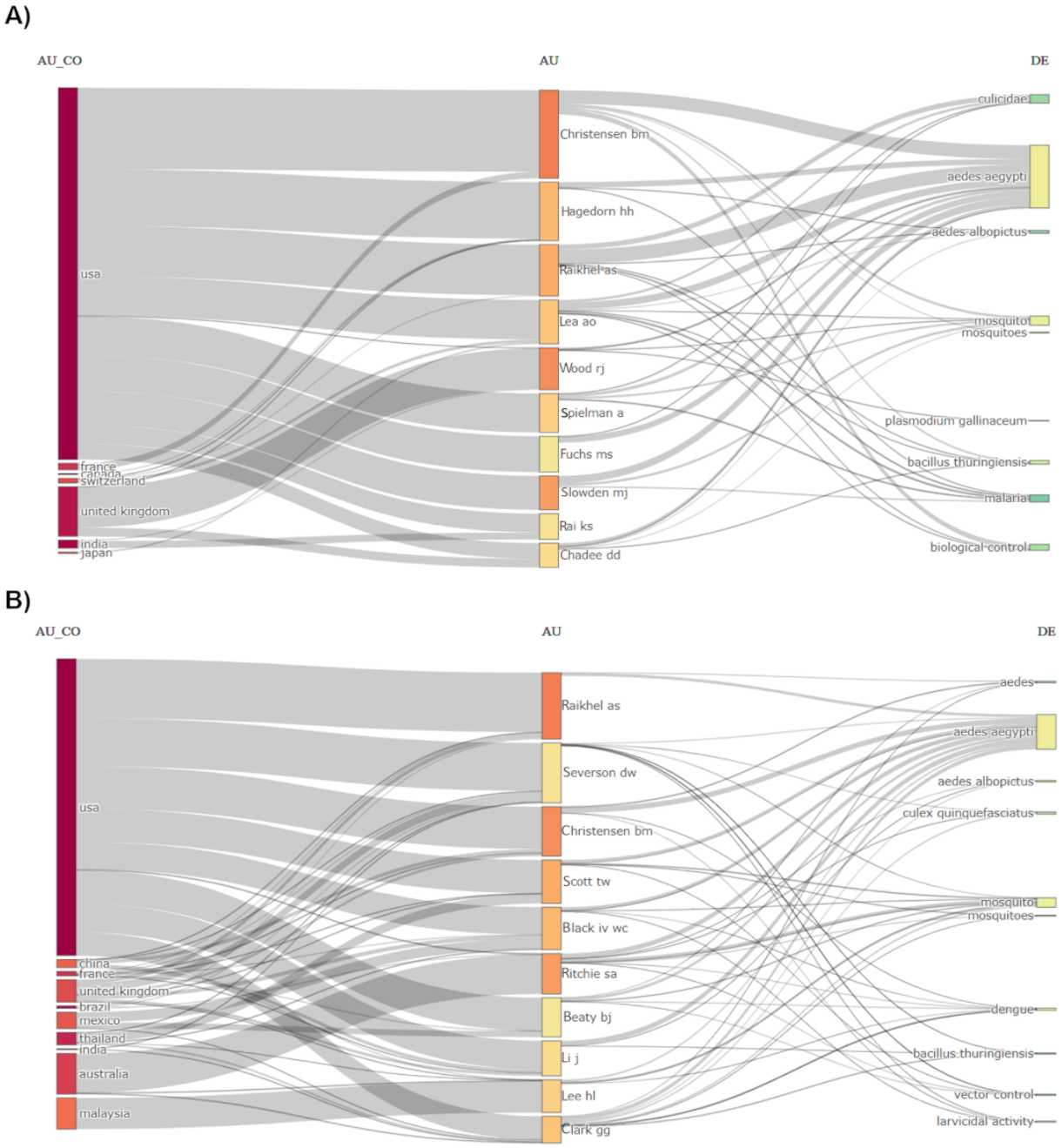


Figure 4 depicts the country of origin of the top authors and their research keywords for each period. In the initial period from 1927-1999, about 90% of the top 10 authors were affiliated with the United States and focused on the study of the Culicidae, *Aedes albopictus*, *Plasmodium gallinaceum*, *Bacillus thuringiensis*, malaria, and biological control (Figure 4A). From 2000 to 2009, more authors from other countries published on *Aedes aegypti*, that was Ritchie SA from Australia and Lee HL from Malaysia. The focus of research has shifted from Malaria to dengue and the additional species that was included was *Culex quinquefasciatus* (Figure 4B). The larvicidal activity was also the focus of the research. From 2010 to 2019, three main authors from Australia topped the productive list, namely Ritchie SA, O'neil SL, and Hoffmann AA, while authors from India contributed many articles, including Benelli G, Murugan K, and Govindarajan M. The top authors focused not only on dengue but also Zika virus (Figure 5A). In the recent period (2020-2023), authors from the United States, Australia, Mexico, China, Brazil, and India were equally productive in publishing their research on *Aedes aegypti* and on similar keywords as the previous periods (Figure 5B).

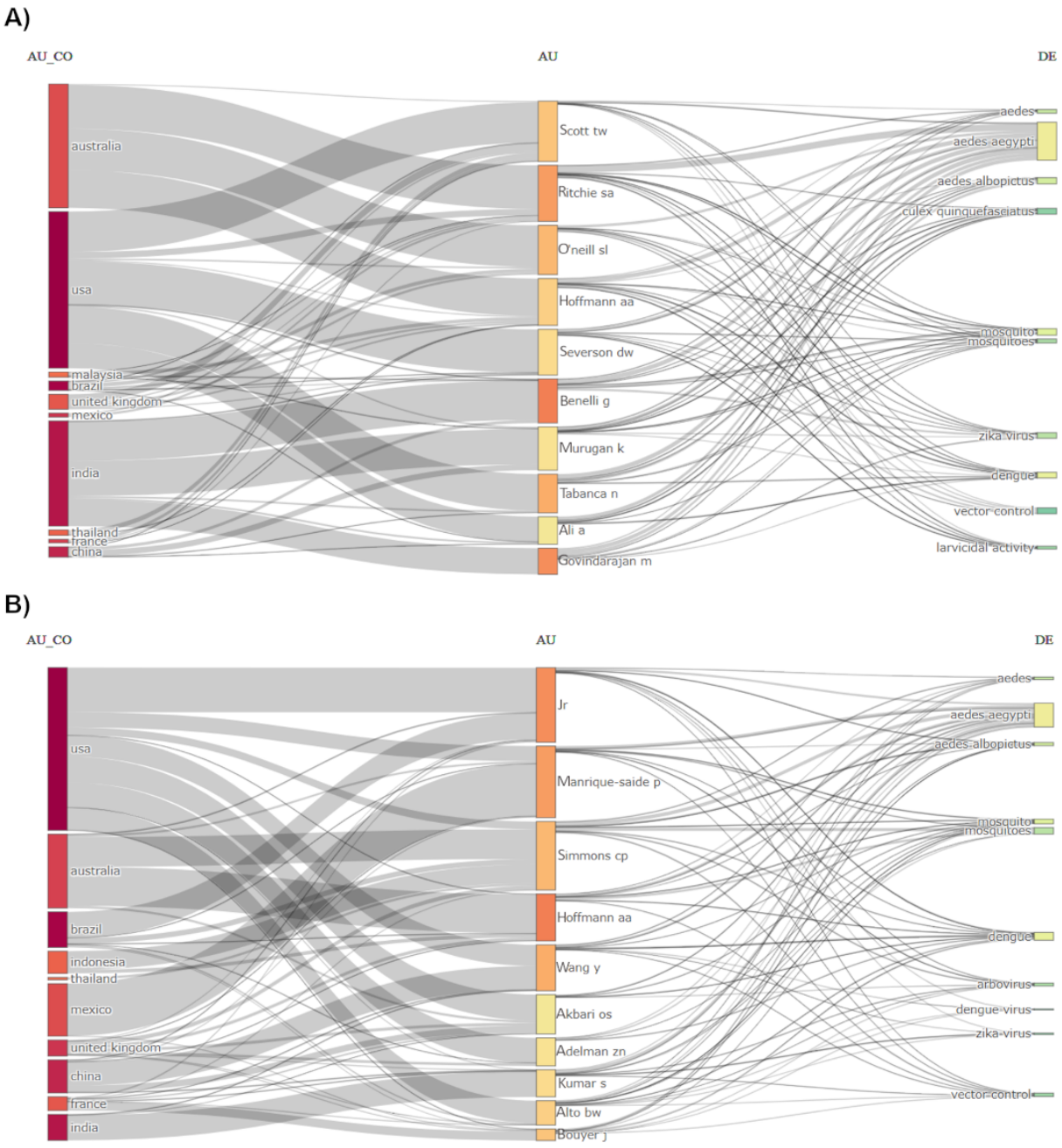
XSL•FO
RenderX

Figure 4. Three-field plot (Sankey diagram) of country, author, and keywords of the cited references for the 10 most productive authors from (A) 1927 to 1999 and (B) 2000 to 2009. AU: authors; AU_CO: countries; DE: keywords.



XSL•FO
RenderX

Figure 5. Three-field plot (Sankey diagram) of country, author, and keywords of the cited references for the 10 most productive authors from (A) 2010 to 2019 and (B) 2020 to 2023. AU: authors; AU_CO: countries; DE: keywords.



Subject Area and Research Trends

Most of the publications were in journals within the subject areas of medicine (7860/30,051; 26.2%); agricultural and biological sciences (5863/30,051; 19.5%); immunology and microbiology (5150/30,051; 17.1%); biochemistry, genetics, and molecular biology (4078/30,051; 13.6%); veterinary science (1513/30,051; 5.0%), and environmental science (1194/30,051; 4.0%). From 2000 to 2019, physics- and astronomy-related publications exhibited the highest growth rate (84), followed by engineering (9.7) and mathematics (8.54).

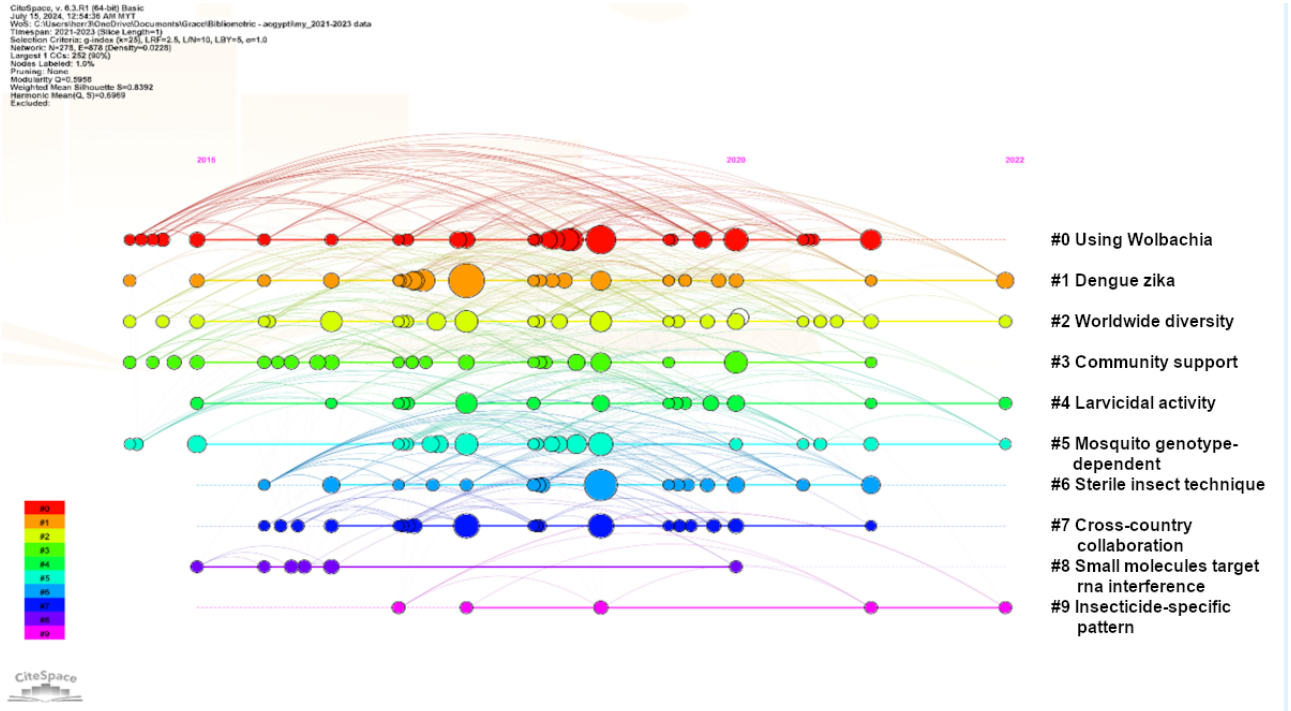
We identified 31 research clusters based on the cocitation reference network. From 2021 to 2023, the temporal evolution of the research topic clusters was discovered with the indication of the cluster silhouette score, size, and mean year of co-cited papers. The research clusters were categorized into 5 major research trends, that was, biological vector control, chemical vector control, and genetic of *Aedes aegypti*, disease-related and collaboration efforts. The main cluster, biological vector control, referred to the largest cluster, #0, concerning "using Wolbachia" (score=0.793; size=42; year=2018) [27] and cluster #6 on "sterile insect technique (SIT)" (score=0.856; size=25; year=2019) [28] (Figure 6). SIT started gaining attention after Wolbachia (#0), as one of the biological vector controls using sterilized male *Aedes aegypti* with either irradiation or exposure to a chemical sterilant in damaging somatic tissue to reduce the mosquito population [29]. The second major research trend concerning disease-related focused on "dengue Zika," #1,

XSL•FO
RenderX

(score=0.839; size=35; year=2019) [19]. The third major research trend was related to collaboration efforts, which referred to "worldwide diversity," #2, (score=0.814; size=33; year=2018) [30], "community support," #3, (Score=0.834; size=29; year=2017) [31] and "cross-country collaboration," #7, (score=0.8; size=24; year=2018) [32]. The research related to chemical vector control contributed to the fourth research trend: "larvicidal activity," #4, (score=0.883; size=26; year=2018) [33] and "insecticide-specific pattern," #9, (score=0.971; size=6; year=2019) [34]. The fifth research trend was related to genetics: "mosquito genotype-dependent," #5, (score=0.877; size=26; year=2018) [35] and "small molecules target RNA interference," #8, (score=0.932; size=6; year=2017) [36].

There were other significant subject areas from the analysis of keywords co-occurrences across 4 periods. Before the year 2000, the focused subject areas included *Bacillus thuringiensis*, electron microscopy, chromosome mapping, virus cultivation, and mosquito control (Figure S1 in Multimedia Appendix 1). From year 2000 to 2009, the research subject area related to *Aedes aegypti* focused on Culicidae, molecular sequence data, insecticides, bacterial toxins, larvae, insecticides, and dengue virus (Figure S2 in Multimedia Appendix 1). The subject areas were then extended to insecticidal activity, insect repellent, bioassay, plant extract, metabolism, genetics, insect proteins, physiology, disease transmission, and population density from the year 2010 to 2020 (Figure S3 in Multimedia Appendix 1). In the recent 4 years from 2020 to 2023, the subject areas focused on insecticide resistance, temefos, pyrethrins, larvicidal activity, metabolism, gene expression, reverse transcription polymerase, prevalence, vector control, arbovirus infections, and Zika virus (Figure S4 in Multimedia Appendix 1). The research focuses on *Aedes aegypti* linking with dengue stayed relatively important throughout the period. The number of publications focused on yellow fever were low and linked to *Aedes aegypti* solely (Figures S5-S8 in Multimedia Appendix 1). The research on chikungunya were not closely related to others research topics in the early period and were only started linking with dengue in the period of 2010-2019 and connecting with Zika in the period of 2020-2023. In 2010-2019, Zika started to become closely related topic with dengue and chikungunya in addition later on.

Figure 6. Visualization of the reference cocitation network time map (Year 2021-2023).



Citation Analysis

The top 10 papers with the highest number of citations are listed in Table 2. The top cited article was "Pathogenesis of dengue: challenges to molecular biology" authored by Halstead SB [37] (n=1355). The second highest cited article was "The global distribution of the arbovirus vectors Aedes aegypti and Aedes albopictus," authored by Kraemer et al [18] (n=1324), followed by the "A Wolbachia symbiont in Aedes aegypti limits infection with dengue, chikungunya, and Plasmodium. Cell. 2009" by Moreira et al [38] (n=1196).

XSL•FO
RenderX

Table 2. Top 10 cited papers.

Article authors and title	Journal	Year	Citation, n (%)	Average citations per year
Halstead SB. Pathogenesis of dengue: challenges to molecular biology [37].	*Science*	1988	1355 (0.28)	38.7
Kraemer MU, Sinka ME, Duda KA, Mylne AQ, Shearer FM, Barker CM, Moore CG, Carvalho RG, Coelho GE, Van Bortel W, Hendrickx G. The global distribution of the arbovirus vectors Aedes aegypti and Aedes albopictus [18].	*eLife*	2015	1324 (0.27)	165.5
Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, Rocha BC, Hall-Mendelin S, Day A, Riegler M, Hugo LE. A Wolbachia symbiont in Aedes aegypti limits infection with dengue, Chikungunya, and Plasmodium [38].	*Cell*	2009	1196 (0.25)	85.4
Gubler DJ. Epidemic dengue and dengue hemorrhagic fever as a public health, social and economic problem in the 21st century [39].	*Trends in Microbiology*	2002	1191 (0.25)	56.7
Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, Aiden EL. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds [40].	*Science*	2017	1136 (0.24)	38.7
Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A single mutation in Chikungunya virus affects vector specificity and epidemic potential [41].	*PLoS Pathogens*	2007	1127 (0.23)	70.4
Musso D, Gubler DJ. Zika virus [42].	*Clinical Microbiology Reviews*	2016	1112 (0.23)	158.9
Weaver SC, Reisen WK. Present and future arboviral threats [43].	*Antiviral Research*	2010	1060 (0.22)	38.7
Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, Muzzi F, Greenfield M, Durkan M, Leong YS, Dong Y, Cook H. Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission [44].	*Nature*	2011	1052 (0.22)	87.7
Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JL, Guimarães KP, Benazzato C, Almeida N, Pignatari GC, Romero S, Polonio CM. The Brazilian Zika virus strain causes birth defects in experimental models [45].	*Nature*	2016	1003 (0.21)	143.3

Journals

The total number of journals that published papers on *Aedes aegypti* from 1928 to 2023 were 160. The leading journal that published articles on the *Aedes aegypti* were *Journal of Medical Entomology* (JME; 758/16,247, 4.7%), *PLOSNeglected Tropical Diseases* (NTD) (604/16,247, 3.7%), *Journal of theAmerican Mosquito Control Association* (AMCA) (505/16,247, 3.1%), *Parasites and Vectors* (493/16,247, 3.0%), and *PLOS One* (401/16,247, 2.5%; Table 3). The journal with the highest number of citations for their publication on *Aedes aegypti* was PLOS NTD (n=25,274), followed by JME (n=23,682) and the *American Journal of Tropical Medicine and Hygiene* (AJTMH; n=19,281; Table 3).

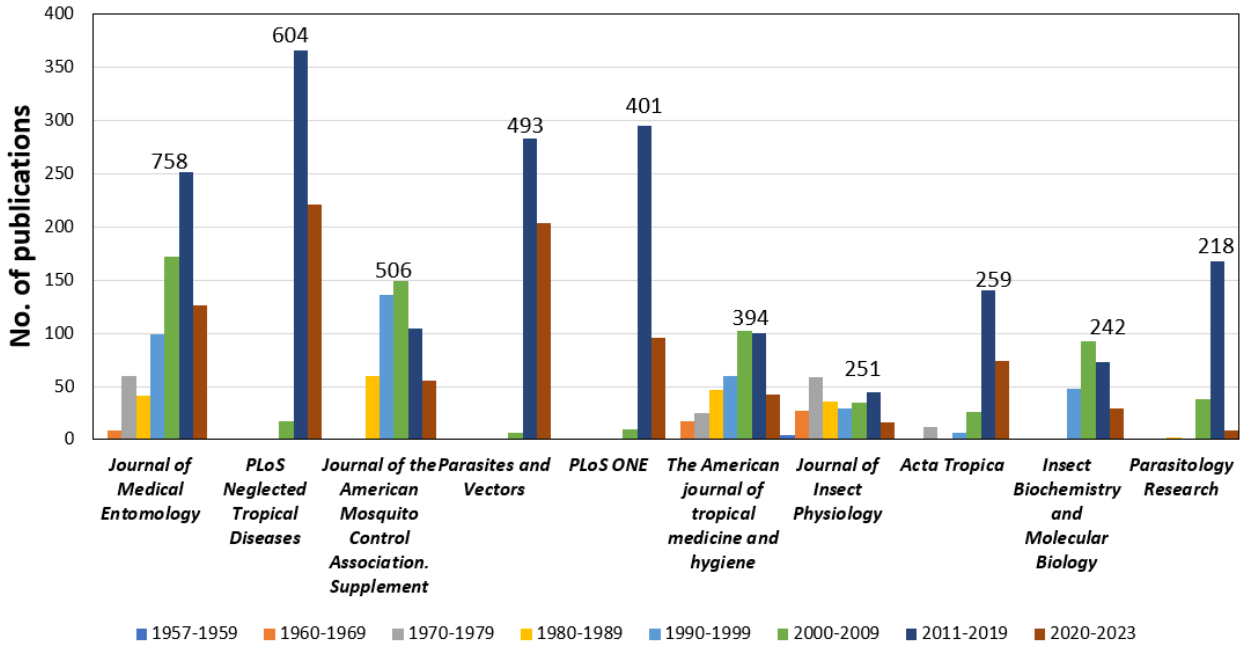
Figure 7 shows the distribution of publications on *Aedes aegypti* by journal across the years from 1957 to 2023. The *Journal of Insect Physiology* was the earliest to publish on *Aedes aegypti*, that was, since 1957, whereas the AJTMH and JME began publishing on *Aedes aegypti* between 1960 and 1970. From 2011 to 2019, most study were published in *PLOS NTD*, *Parasites and Vectors*, *PLOS One*, *Parasitology Research*, and *Acta Tropica*.

Table 3. Top 10 journals by number of papers on Aedes aegypti (from 1927-2023) and 5-year impact factor (2019-2023).

Journal	Number of papers (N=16,247), n (%)	Total citations (N=481,479), n (%)	5-year impact factor in 2023
Journal of Medical Entomology	758 (4.67)	23,682 (4.92)	2
PLOS [a] *neglected tropical diseases*	604 (3.72)	25,274 (5.25)	3.6
Journal of the American Mosquito Control Association	505 (3.11)	10,818 (2.25)	1
Parasites and Vectors	493 (3.03)	11,896 (2.47)	3.3
PLOS One	401 (2.47)	14,877 (3.09)	3.3
American Journal of Tropical Medicine and Hygiene	357 (2.2)	19,281 (4.00)	2.2
Acta Tropica	259 (1.59)	6312 (1.31)	2.4
Journal of Insect Physiology	251 (1.54)	8193 (1.70)	2.3
Insect Biochemistry and Molecular Biology	242 (1.49)	11,658 (2.42)	3.7
Parasitology Research	218 (1.34)	11,767 (2.44)	1.9

XSL•FO
RenderX

Figure 7. Number of publications by the top 10 journals by decade.



Country Analysis

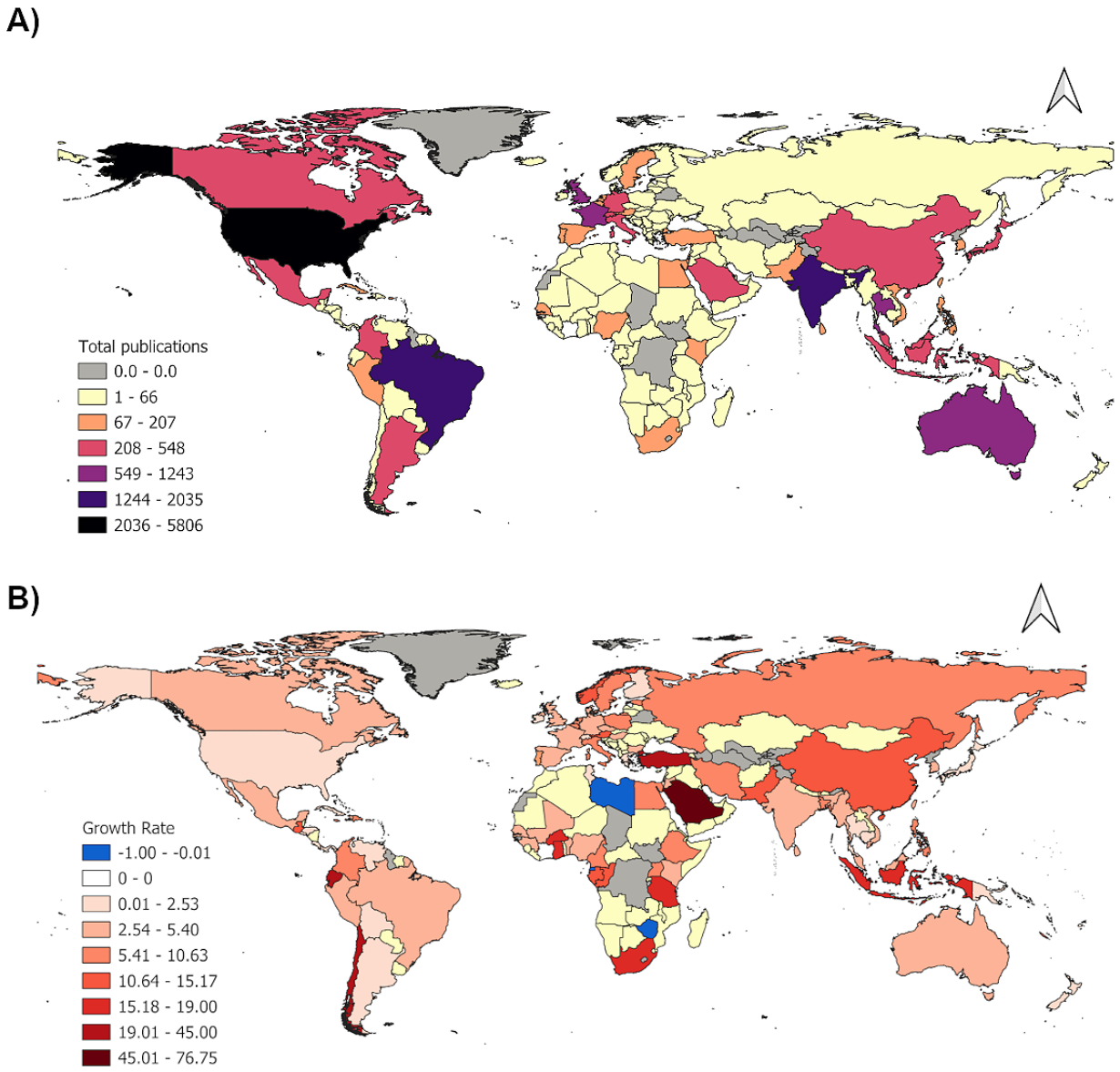
More than one-third of the publications during the study period were by authors from the United States (5806/23,538, 24.7%), Brazil (2035/23,538, 8.6%), and India (1839/23,538, 7.8%). The next top 7 countries were from Europe (United Kingdom: 1243/23,538, 5.3%; and France: 898/23,538, 3.8%), Asia (Thailand: 678/23,538, 2.9%; Malaysia: 455/23,538, 1.9%; and China: 541/23,538, 2.3%), Mexico (548/23,538, 2.3%), and Australia (779/23,538, 3.3%). The decadal growth rate (from 2000 to 2019) of the articles by country was topped by Saudi Arabia (76.75), followed by Turkey (45), Chile (29), Ecuador (24), and Ghana (19) (Figure 8A). High decadal growth rates were also observed for Africa (Burkina Faso: 17.67; South Africa: 16.33; and the United Republic of Tanzania: 16), Asia (Hong Kong: 18; Indonesia: 16.6; and Pakistan: 14.5), Europe (Austria: 15.17; Norway: 13.5; and Estonia: 9), South America (Colombia: 9.52; French Guiana: 5.4; and Brazil: 4.82), Australia (4.17), and the United States (2.46; Figure 8B).

Coauthorship analysis of 131 countries revealed that the majority of coauthors were authors from the United States who collaborated with authors from Brazil and South America (Figure 9). Authors from India collaborated with Australian, Taiwanese, and South Korean authors. UK authors mainly collaborated with Thai, Malaysian, and Sri Lankan authors. Authors in France collaborated primarily with countries in Africa including Uganda, Congo, Cameron, Senegal, and Ghana. The European authors, that was, in Germany, Italy, and Spain worked closely with the collaborators from Saudi Arabia, Bangladesh, Egypt, and Turkey.

The country with the highest total citations was the United States (182,414/481,479, 37.89%), which was 5 times higher than the second highest country, Brazil (39,115/481,479, 8.12%). Switzerland (7469/481,479, 1.55%) and Italy (7,239/481,479, 1.50%) were both in the top 10 highest number of citations, even though they had low numbers of articles (Switzerland: n=302; and Italy: n=307). The rest of the countries in the top 10 list of total citation based on origin of corresponding author were India (32,646/481,479, 6.78%), United Kingdom (30,328/481,479, 6.30%), Australia (25,242/481,479, 5.24%), France (23,626/481,479, 4.91%), Thailand (12,154/481,479, 2.52%), and Mexico (6450/481,479, 1.34%).

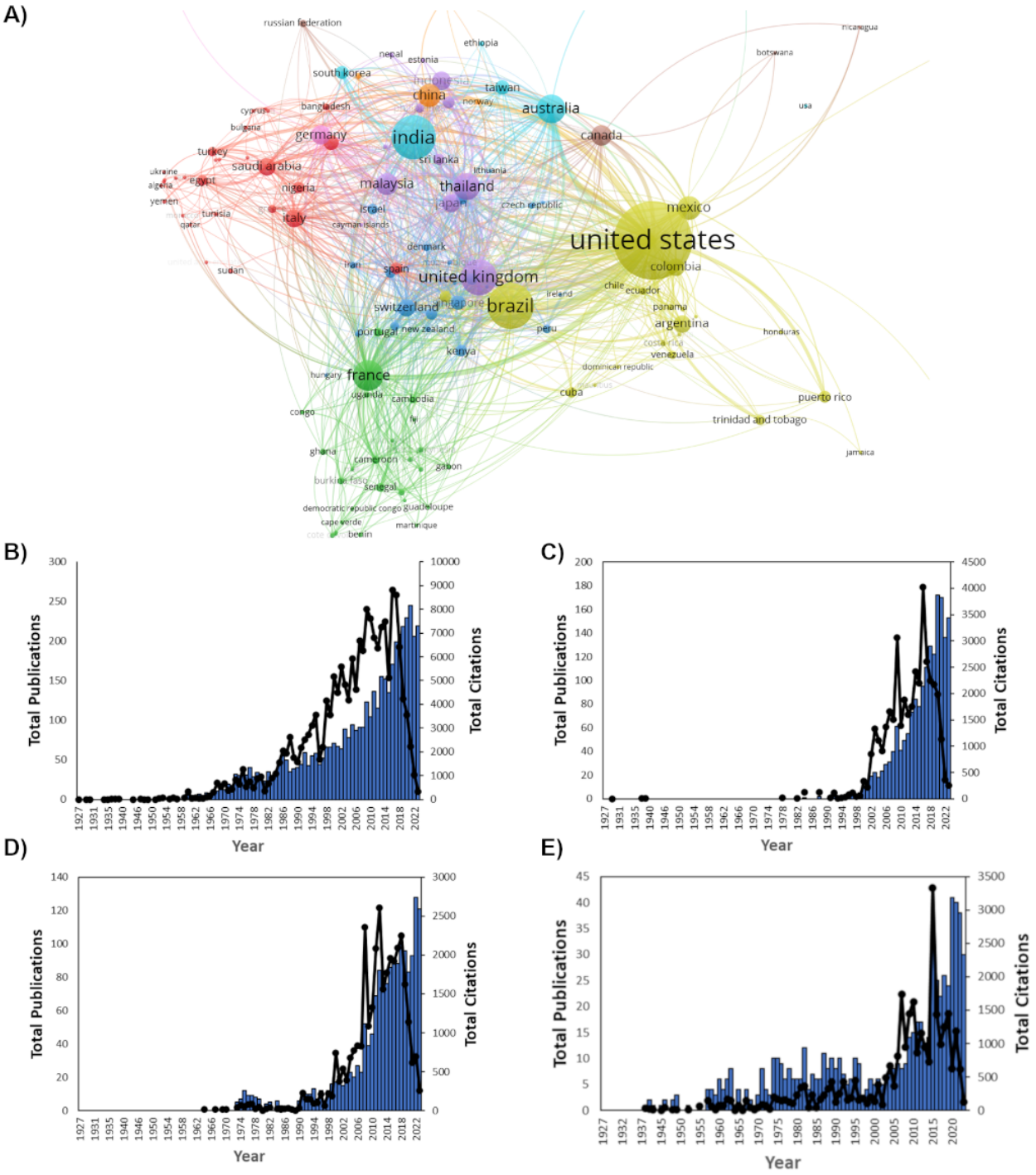
XSL•FO
RenderX

Figure 8. A map of the (A) total publications of Aedes aegypti by country for the years 1927-2023 and (B) growth rate of the scientific articles on Aedes aegypti by country for the year 2000-2009 to 2010-2023.



XSL•FO
RenderX

Figure 9. Co-authorship analysis of 131 countries. (A) Co-authorship by country: publications and total citations in the top 4 countries including (B) the United States, (C) Brazil, (D) India, and (E) the United Kingdom.



Discussion

Principal Findings

The medically important mosquitoes, in particular, *Aedes aegypti*, has been studied for centuries. Female *Aedes aegypti* was one of the main vectors for transmitting viruses, including dengue virus, Zika virus, chikungunya virus, and yellow fever virus. In this study, the detailed evaluation of the *Aedes aegypti* published works informed the current knowledge gaps and future direction of the control of the vector. During the study period, 16,247 items were published and cited 481,479 times. In the recent 3 years, the literature focused on 5 areas: biological vector control studies, chemical vector control studies, genetic studies, disease-related studies, and collaboration efforts. The minor changes in the research trend were observed when comparing Vega-Almeida et al [12] (year 2006-2015) with the research focus on epidemiology, gene expression and biological control, larvicidal and insecticidal effects, and reproduction and insecticide resistance. The biological vector control using Wolbachia has been gaining more attention as many countries proved its efficiency in reducing dengue cases [44,46,47]. The

XSL•FO
RenderX

ongoing Sterile insect techniques pilot projects in many tropical and sub-tropical countries have shown the effectiveness of using irradiated mosquitoes in reducing the wild population [48]. The research direction will be the biological vector control projects at a larger scale or national operational level in the future. Lately, dengue, Zika, and chikungunya were emerging research topics related to *Aedes aegypti*. In recent years, there has been a noticeable rise in the number of articles addressing *Aedes aegypti* in discussing the effectiveness of vector control strategies. One of the control measures of the effectiveness of vector control strategies was the insecticide resistance of *Aedes aegypti*. Insecticides remain widely used in vector control interventions resulted in mosquito resistance to insecticides in 4 classes, including organophosphates, carbamates, pyrethroids and the organochlorine dichlorodiphenyltrichloroethane [49], and more in recent years. The insecticide resistance level of *Aedes aegypti* were conducted with various methods, for example, the WHO (World Health Organization) tube test [50], WHO bottle bioassay [51-53], the Centers for Disease Control and Prevention's bottle bioassay [54,55]. The WHO showed capital interest in the insecticide resistance study on malaria and dengue vector mosquitoes that several guidelines were provided to discuss the monitoring of insecticide resistance in mosquito vectors [56,57].

Genetic studies in relation to the gene knockdown of insecticides and the genomic profiles and relationship of the vector and patient were also the main research focus in recent years as the knowledge of genes requires more exploration and elucidation. The detailed genome mapping, RNA-seq data alignment, and gene expression quantification methods were investigated in detail to understand the relationship between mosquito genotype and the microbes in the science of vector mosquito genetic control [35,58,59]. Zika is still without cure and the symptoms could be severe, such as Guillain-Barré syndrome, congenital malformations in infancy, neuropathy, and myelitis in adults and older children. Many current studies focused on the vector competence of *Aedes aegypti*, particularly the Zika virus strain' infectivity, dissemination, and transmission rates as well as the influence of the larval microbiome on mosquito genotype-dependent [35,60]. Understanding the antiviral pathway in using the small molecule RNA interference and jak or stat signaling is crucial in controlling Zika virus infection in *Aedes aegypti* [36,61].

The booms of publications on *Aedes aegypti* after the year 2000 may be due to the increase in dengue cases globally, from 505,430 to 6.5 million from the year 2000 to 2023 [62,63]. Since 2007, there have been several Zika outbreaks in Africa, the Americas, Asia, and the Pacific. Zika was briefly declared as a Public Health Emergency of International Concern in 2016 [64]. The threat of chikungunya outbreaks in recent decades was unprecedented, especially in 2013-14 in the Caribbean and Latin America [65]. Moreover, the WHO declared the yellow fever as a global epidemic threat due to international travel [65]. The number of publications on *Aedes aegypti* and other vector is expected to increase in the future if infectious diseases caused by this vector remain uncontrolled or are exacerbated by climate change, rapid urbanization, and international human mobility. In the recent 2 decades, there was a surge of the growth rate of the publication on *Aedes aegypti*. The dengue endemic countries in the region of Western Asia, South America, and Western and Southern Africa produced more research output focus on the *Aedes aegypti*. However, countries in the northern and central Africa have been underrepresented in the research studies on *Aedes aegypti*. Studies in Burkina Faso, Ghana, South Africa, and Tanzania-countries with higher growth rate in publications on *Aedes aegypti*–can serve as the valuable references for other African nations with similar human environment in the study of origin, biology, behaviors, habitat, insecticide sensitivity of vectors, etc [66-70].

The analysis of coauthorship revealed that publications on *Aedes aegypti* were primarily collaborative efforts involving multiple authors from various specialized research fields. Authors from the United States constituted 45% of the top 20 most productive authors. This was likely due to increased access to funding resources. The top 20 productive authors originated from Australia, Asia (Thailand, India, and Malaysia), South America (Brazil and Trinidad and Tobago), and Europe (France and Italy). Authors who published on diseases transmitted by *Aedes aegypti* that are of global relevance, such as Gubler DJ and Halstead SB, had high numbers of citations and co-citations. During the Zika outbreak, authors who published on *Aedes aegypti* received more citations because of the increased public health interest in disease patterns, transmission links, and vector control methods related to Zika [42]. Often, authors gained more citations after years of experience in disseminating research findings [71].

The co-citation analysis provided a deeper understanding of vital research trends on *Aedes aegypti*. Gubler DJ was the highest co-cited author but was not among the top 20 productive authors. Gubler (H-index 94) was central in bridging researchers in the field for his significant works on dengue, antibody-dependent enhancement, flavivirus, Zika virus, microcephaly, and yellow fever. Gubler's review papers on dengue and dengue hemorrhagic fevers globally and locally were cited by the majority of dengue research papers, especially in the introduction sections [72]. The top co-cited article also highlighted a global study similar to Gubler DJ whereas Bhatt et al (2013) mapped and discussed the global distribution of the burden of dengue. Co-citation analysis can answer the research question, "Who are the central, peripheral, and bridging researchers in the field, and how has the structure developed over time?" [73] by connecting journals, authors, and various documents [74] regarding the intellective structure. The cocitation cluster patterns provide a new way to study the specialty structure of science [75]. Further studies on the cocitation trends by the years can provide insight and understanding of how the research developed over time.

There appeared to be links between authorship, citation impacts, collaborations, and funding. Higher funding support is strongly associated with higher citation impacts [76]. Funders were acknowledged in only 52.7% of the articles, based on the available information that showed the United States contribute to most of the top cited authors and articles on *Aedes aegypti*. The United States together with Brazil published the most study on *Aedes aegypti* aligning with other worldwide dengue bibliometric studies, showing these countries were the most

XSL•FO
RenderX

productive countries [77], given that *Aedes aegypti* is the main vector for dengue disease in the countries. Authors from neighboring countries tended to collaborate more within their regions. International coauthorship has a strong positive effect on the number of articles and citations a country produces [78,79]. This regional clustering is evident in Figure 8A, where the United States collaborated with Mexico, Colombia, Panama, and Argentina, while India worked with Sri Lanka, Thailand, and Bangladesh. Recently, China has shown a steep increase in publications on *Aedes aegypti* due to increased research funding and higher research focus on infectious disease [80-82].

Journal analysis provided information on the journals with the highest impact articles related to *Aedes aegypti* worldwide. The top journal, JME, a bimonthly publication focused on medical entomology and medical acarology, especially arthropods of public health importance, has published articles on *Aedes aegypti* since 1957 and exhibited an increasing trend in the number of articles for the past 60 years. JME's impact factor was 2 (The year 2019-2023) and was the second highest in total citations. Although PLOS NTD contributed the second highest number of publications on *Aedes aegypti*, it obtained the highest number of citations since its first publication on *Aedes aegypti* in 2008 and with a recent impact factor of 3.6 (for the year 2019-2023), indicating it has a great impact. PLOS NTD was also reported as the leading journal in publications on Leishmania [83] and neglected tropical diseases [84] in the Latin Americas and the Caribbean. However, these journals were grouped into different clusters as JME published more articles on *Aedes aegypti* physiology and vector control, whereas PLOS NTD articles covered more virus and genetic studies. The 10 leading journals mostly fell in the categories of agricultural and biological sciences, immunology and microbiology, and medicine of Clarivate's Journal Citation Report. Most of the journals that fell into multiple subject categories gained higher citations than those in single or dual categories [85].

The limitation of the current study should not be omitted. The study only sourced the data from Scopus due to the difficulties in combining the control measure across multiple databases, hence it may not capture all the papers from other databases such as Web of Science, Google Scholar, and Microsoft Academic. Besides, the qualitative components of the research were inaccessible in terms of peer review, ethical concerns and societal influence. Outputs from other academic discipline such as patents, produced systems developed and widely used, policy papers, white papers, reports produced for government and other public organizations, and exhibition were not included. In short, this study remains as valuable resources for comprehending the direction and significance of research.

Conclusions

In conclusion, the research on *Aedes aegypti* increased at a relative growth rate of 12.1 especially in the year 2000 to 2009, with a total number of 16,247 articles in 160 journals and were cited for 481,479 times. The prolific authors were mainly from the United States and the top co-cited authors across the years were elucidated. The coauthorship was collaborative involving multiple authors from various specialized research fields and countries, especially neighboring countries. Central and bridging researchers of the study on *Aedes aegypti* were identified and discussed based on the co-citation analysis, and this will provide better understanding how the research changed over time. The researchers' next area of interest may be the research direction. The leading journal were JME, PLOS NTD, AMCA, *Parasites and Vectors*, and *PLOS One*, with the top cited article titled "Pathogenesis of dengue: challenges to molecular biology" by Halstead SB in *Science*. Leading countries in published articles such as the United States and Brazil were also countries that funded higher number of published articles on *Aedes aegypti*.

The strengths of the study include the capability to identify the recent 3 years' research trends: "using Wolbachia," "Dengue Zika," "worldwide diversity," "community support," "larvicidal activity," "mosquito genotype-dependent," "sterile insect technique," "cross-country collaboration," "small molecules target RNA interference," and "insecticide-specific pattern." The recent research mainly focused on biological control, that is, an alternative to chemical control. This will be awareness for the scientific communities on the need to study the medicinal important vector, in which the researchers or entomologists could understand the current knowledge gap on *Aedes aegypti* and to plan for future research pathways. This study also contributed to the public health stakeholder and funder on the current research direction and knowledge gap for better decision-making and research priorities in determining the suitable intervention for vector control.

Acknowledgments

We would like to thank the Director General of Health Malaysia and the Director of the Institute for Medical Research (IMR) for their permission to publish this article. This work was accomplished under the project of the National Institute of Health, Ministry of Health Malaysia (NMRR-22-01333-RCQ).

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

YLC handled the conceptualization, data curation, formal analysis, methodology, visualization, and writing of the original draft. KHL worked on the conceptualization, methodology, supervision, visualization, review and editing of the writing. BSG worked

XSL•FO
RenderX

on the supervision, visualization, review and editing of the writing. MHMH, MKCI, AA, TYT and HLL worked on the data curation and formal analysis. SMG and YKC did the methodology, review and editing of the writing.

Conflicts of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Multimedia Appendix 1
Co-occurrence network of keywords.
[DOCX File , 5397 KB - ijmr_v14i1e65844_app1.docx]

References

1. Clements AN, Harbach RE. Controversies over the scientific name of the principal mosquito vector of yellow fever virus - expediency versus validity. J Vector Ecol 2018;43(1):1-14 [FREE Full text] [doi: 10.1111/jvec.12277] [Medline: 29757506]
2. Edwards FW, Patton WS. The identity of L. Ann Trop Med Parasitol 1933;27(2):182-184. [doi: 10.1080/00034983.1933.11684765]
3. Weetman D, Kamgang B, Badolo A, Moyes C, Shearer F, Coulibaly M, et al. Aedes mosquitoes and aedes-borne arboviruses in Africa: current and future threats. Int J Environ Res Public Health 2018;15(2):220 [FREE Full text] [doi: 10.3390/ijerph15020220] [Medline: 29382107]
4. Tedjou AN, Kamgang B, Yougang AP, Njiokou F, Wondji CS. Update on the geographical distribution and prevalence of aedes aegypti and aedes albopictus (Diptera: Culicidae), two major arbovirus vectors in cameroon. PLoS Negl Trop Dis 2019;13(3):e0007137 [FREE Full text] [doi: 10.1371/journal.pntd.0007137] [Medline: 30883552]
5. Nordin O, Ney TG, Ahmad NW, Benjamin S, Lim LH. Identification of aedes aegypti (L.) and aedes albopictus (Skuse) breeding habitats in dengue endemic sites in kuala lumpur federal territory and selangor state, malaysia. Southeast Asian J Trop Med Publ Hlth 2017;48:786-798.
6. Leta S, Beyene TJ, De Clercq EM, Amenu K, Kraemer MU, Revie CW. Global risk mapping for major diseases transmitted by aedes aegypti and aedes albopictus. Int J Infect Dis 2018;67:25-35 [FREE Full text] [doi: 10.1016/j.ijid.2017.11.026] [Medline: 29196275]
7. Ivancheva L. Scientometrics today: a methodological overview. COLLNET J Sci Inf Manag 2008;2(2):47-56. [doi: 10.1080/09737766.2008.10700853]
8. Han Y, Wennersten SA, Lam MPY. Working the literature harder: what can text mining and bibliometric analysis reveal? Expert Rev Proteomics 2019;16(11-12):871-873 [FREE Full text] [doi: 10.1080/14789450.2019.1703678] [Medline: 31822126]
9. Santos SR, Vianna M. Scientometric analysis of fisheries science studies of western atlantic species of (Paralichthyidae: Pleuronectiformes). Reviews in Fisheries Science & Aquaculture 2018;26(4):443-459. [doi: 10.1080/23308249.2018.1452896]
10. Araújo WSD. 30 years of research on insect galls in Brazil: a scientometric review. Pap. Avulsos Zool 2018;58:e20185834. [doi: 10.11606/1807-0205/2018.58.34]
11. Miguel TB, Calvão LB, Vital MVC, Juen L. A scientometric study of the order odonata with special attention to Brazil. International Journal of Odonatology 2017;20(1):27-42. [doi: 10.1080/13887890.2017.1286267]
12. Vega-Almeida RL, Carrillo-Calvet H, Arencibia-Jorge R. Diseases and vector: a 10 years view of scientific literature on aedes aegypti. Scientometrics 2018;115(3):1627-1634. [doi: 10.1007/s11192-018-2650-9]
13. Halim NMHNA, Dom NC, Abdullah S, Salim H. A bibliometric analysis of dengue vector-related publications in the science citation index expanded. Malaysian J Med Heal Sci 2021;17(3):22.
14. Piovezan-Borges AC, Valente-Neto F, Urbieta GL, Laurence SGW, de Oliveira Roque F. Global trends in research on the effects of climate change on aedes aegypti: international collaboration has increased, but some critical countries lag behind. Parasit Vectors 2022;15(1):346 [FREE Full text] [doi: 10.1186/s13071-022-05473-7] [Medline: 36175962]
15. Ammed K, Gupta R, Tiwari R. Dengue research in India: a scientometric analysis of publications, 2003-12. Int J Med Public Health 2014;4(1):1. [doi: 10.4103/2230-8598.127114]
16. Singh N. Scientometric analysis of research on Zika virus. Virusdisease 2016;27(3):303-306 [FREE Full text] [doi: 10.1007/s13337-016-0339-3] [Medline: 28466043]
17. Sindhania A, Baruah K, Katewa A, Sharma YP. Tracing the trajectory of and research: eight decades of bibliometric retrospect. Vector Borne Zoonotic Dis 2025;25(3):155-166. [doi: 10.1089/vbz.2024.0081] [Medline: 39585388]
18. Kraemer MUG, Sinka ME, Duda KA, Mylne AQN, Shearer FM, Barker CM, et al. The global distribution of the arbovirus vectors aedes aegypti and ae. albopictus. Elife 2015;4:e08347 [FREE Full text] [doi: 10.7554/eLife.08347] [Medline: 26126267]
19. Delrieu M, Martinet JP, O'Connor O, Viennet E, Menkes C, Burtet-Sarramegna V, et al. Temperature and transmission of chikungunya, dengue, and zika viruses: a systematic review of experimental studies on and. Curr Res Parasitol Vector Borne Dis 2023;4:100139. [doi: 10.1016/j.crpvbd.2023.100139] [Medline: 37719233]

XSL•FO
RenderX

20. Oliveira OJ. de, ed. Bibliometrics. Brazil: Sao Paulo State University - UNESP; 2024.
21. Delpeuch A, Morris T, Huynh D, Mazzocchi S, Guidry T, Stephens O, et al. OpenRefine/OpenRefine: 3.8.2. Published online July 2024. URL: https://zenodo.org/records/12689605 [accessed 2024-07-08]
22. Waltman L, van Eck NJ, Noyons ECM. A unified approach to mapping and clustering of bibliometric networks. Journal of Informetrics 2010;4(4):629-635. [doi: 10.1016/j.joi.2010.07.002]
23. Aria M, Cuccurullo C. bibliometrix : an r-tool for comprehensive science mapping analysis. Journal of Informetrics 2017;11(4):959-975 [FREE Full text] [doi: 10.1016/j.joi.2017.08.007]
24. Small H. Co-citation context analysis and the structure of paradigms. J Doc 1980;36(3):183-196. [doi: 10.1108/eb026695]
25. Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J. Am. Soc. Inf. Sci 2005;57(3):359-377. [doi: 10.1002/asi.20317]
26. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53-65 [FREE Full text]
27. Edenborough KM, Flores HA, Simmons CP, Fraser JE. Correction for edenborough et al., "Using to Eliminate Dengue: Will the Virus Fight Back?". J Virol 2021;95(18):e0095321 [FREE Full text] [doi: 10.1128/JVI.00953-21] [Medline: 34432527]
28. Marina CF, Liedo P, Bond JG, R Osorio A, Valle J, Angulo-Kladt R, et al. Comparison of ground release and drone-mediated aerial release of sterile males in southern mexico: efficacy and challenges. Insects 2022;13(4):347 [FREE Full text] [doi: 10.3390/insects13040347] [Medline: 35447790]
29. Aldridge RL, Gibson S, Linthicum KJ. Aedes aegypti controls AE. aegypti: SIT and IIT-An overview. J Am Mosq Control Assoc 2024;40(1):32-49 [FREE Full text] [doi: 10.2987/23-7154] [Medline: 38427588]
30. Misbah-Ul-Haq M, Carvalho DO, Duran De La Fuente L, Augustinos AA, Bourtzis K. Genetic stability and fitness of red-eye genetic sexing strains with pakistani genomic background for sterile insect technique applications. Front Bioeng Biotechnol 2022;10:871703 [FREE Full text] [doi: 10.3389/fbioe.2022.871703] [Medline: 35433649]
31. Sánchez-González L, Adams LE, Saavedra R, Little EM, Medina NA, Major CG, et al. Assessment of community support for wolbachia-mediated population suppression as a control method for aedes aegypti mosquitoes in a community cohort in puerto rico. PLoS Negl Trop Dis 2021;15(12):e0009966 [FREE Full text] [doi: 10.1371/journal.pntd.0009966] [Medline: 34871301]
32. Longbottom J, Walekhwa AW, Mwingira V, Kijanga O, Mramba F, Lord JS. Aedes albopictus invasion across Africa: the time is now for cross-country collaboration and control. Lancet Glob Health 2023;11(4):e623-e628 [FREE Full text] [doi: 10.1016/S2214-109X(23)00046-3] [Medline: 36841255]
33. Leandro ADS, Ayala MJC, Lopes RD, Martins CA, Maciel-de-Freitas R, Villela DAM. Entomo-virological surveillance applied for prediction of dengue transmission: a spatio-temporal modeling study. Pathogens 2022;12(1):4 [FREE Full text] [doi: 10.3390/pathogens12010004] [Medline: 36678352]
34. Derilus D, Impoinvil LM, Muturi EJ, McAllister J, Kenney J, Massey SE, et al. Comparative transcriptomic analysis of insecticide-resistant from puerto rico reveals insecticide-specific patterns of gene expression. Genes (Basel) 2023;14(8):1626 [FREE Full text] [doi: 10.3390/genes14081626] [Medline: 37628677]
35. Accoti A, Multini LC, Diouf B, Becker M, Vulcan J, Sylla M, et al. The influence of the larval microbiome on susceptibility to zika virus is mosquito genotype-dependent. PLoS Pathog 2023;19(10):e1011727 [FREE Full text] [doi: 10.1371/journal.ppat.1011727] [Medline: 37903174]
36. Trammell CE, Ramirez G, Sanchez-Vargas I, St Clair LA, Ratnayake OC, Luckhart S, et al. Coupled small molecules target RNA interference and JAK/STAT signaling to reduce zika virus infection in aedes aegypti. PLoS Pathog 2022;18(4):e1010411 [FREE Full text] [doi: 10.1371/journal.ppat.1010411] [Medline: 35377915]
37. Halstead SB. Pathogenesis of dengue: challenges to molecular biology. Science 1988;239(4839):476-481. [doi: 10.1126/science.239.4839.476]
38. Moreira LA, Iturbe-Ormaetxe I, Jeffery JA, Lu G, Pyke AT, Hedges LM, et al. A wolbachia symbiont in aedes aegypti limits infection with dengue, chikungunya, and plasmodium. Cell 2009;139(7):1268-1278. [doi: 10.1016/j.cell.2009.11.042] [Medline: 20064373]
39. Gubler DJ. Epidemic dengue/dengue hemorrhagic fever as a public health, social and economic problem in the 21st century. Trends Microbiol 2002;10(2):100-103. [doi: 10.1016/s0966-842x(01)02288-0] [Medline: 11827812]
40. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the genome using Hi-C yields chromosome-length scaffolds. Science 2017;356(6333):92-95 [FREE Full text] [doi: 10.1126/science.aal3327] [Medline: 28336562]
41. Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A single mutation in chikungunya virus affects vector specificity and epidemic potential. PLoS Pathog 2007;3(12):e201 [FREE Full text] [doi: 10.1371/journal.ppat.0030201] [Medline: 18069894]
42. Musso D, Gubler DJ. Zika virus. Clin Microbiol Rev 2016;29(3):487-524 [FREE Full text] [doi: 10.1128/CMR.00072-15] [Medline: 27029595]
43. Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res 2010;85(2):328-345 [FREE Full text] [doi: 10.1016/j.antiviral.2009.10.008] [Medline: 19857523]

XSL•FO
RenderX

44. Hoffmann AA, Montgomery BL, Popovici J, Iturbe-Ormaetxe I, Johnson PH, Muzzi F, et al. Successful establishment of wolbachia in aedes populations to suppress dengue transmission. Nature 2011;476(7361):454-457. [doi: 10.1038/nature10356] [Medline: 21866160]
45. Cugola FR, Fernandes IR, Russo FB, Freitas BC, Dias JLM, Guimarães KP, et al. The Brazilian Zika virus strain causes birth defects in experimental models. Nature 2016;534(7606):267-271 [FREE Full text] [doi: 10.1038/nature18296] [Medline: 27279226]
46. Nazni WA, Hoffmann AA, NoorAfizah A, Cheong YL, Mancini MV, Golding N, et al. Establishment of wolbachia strain wAlbB in malaysian populations of aedes aegypti for dengue control. Curr Biol 2019;29(24):4241-4248.e5 [FREE Full text] [doi: 10.1016/j.cub.2019.11.007] [Medline: 31761702]
47. Pinto SB, Riback TIS, Sylvestre G, Costa G, Peixoto J, Dias FBS, et al. Effectiveness of wolbachia-infected mosquito deployments in reducing the incidence of dengue and other aedes-borne diseases in niterói, Brazil: a quasi-experimental study. PLoS Negl Trop Dis 2021;15(7):e0009556. [doi: 10.1371/journal.pntd.0009556] [Medline: 34252106]
48. Guidance framework for testing the sterile insect technique as a vector control tool against aedes-borne diseases. (Yadav RS, ed.). World Health Organization. WHO Team C of NTD (NTD). 2020. URL: https://tinyurl.com/5d55vsm5 [accessed 2020-04-26]
49. Moyes CL, Vontas J, Martins AJ, Ng LC, Koou SY, Dusfour I, et al. Contemporary status of insecticide resistance in the major aedes vectors of arboviruses infecting humans. PLoS Negl Trop Dis 2017;11(7):e0005625 [FREE Full text] [doi: 10.1371/journal.pntd.0005625] [Medline: 28727779]
50. Su X, Guo Y, Deng J, Xu J, Zhou G, Zhou T, et al. Fast emerging insecticide resistance in aedes albopictus in guangzhou, China: alarm to the dengue epidemic. PLoS Negl Trop Dis 2019;13(9):e0007665 [FREE Full text] [doi: 10.1371/journal.pntd.0007665] [Medline: 31525199]
51. Badolo A, Sombié A, Pignatelli PM, Sanon A, Yaméogo F, Wangrawa DW, et al. Insecticide resistance levels and mechanisms in aedes aegypti populations in and around ouagadougou, burkina faso. PLoS Negl Trop Dis 2019;13(5):e0007439 [FREE Full text] [doi: 10.1371/journal.pntd.0007439] [Medline: 31120874]
52. Namountougou M, Soma DD, Balboné M, Kaboré DA, Kientega M, Hien A, et al. Monitoring insecticide susceptibility in populations from the two biggest cities, ouagadougou and bobo-dioulasso, in burkina faso: implication of metabolic resistance. Trop Med Infect Dis 2020;5(2):84 [FREE Full text] [doi: 10.3390/tropicalmed5020084] [Medline: 32471266]
53. Rasli R, Cheong YL, Che Ibrahim MK, Farahininajua Fikri SF, Norzali RN, Nazarudin NA, et al. Insecticide resistance in dengue vectors from hotspots in Selangor, Malaysia. PLoS Negl Trop Dis 2021;15(3):e0009205 [FREE Full text] [doi: 10.1371/journal.pntd.0009205] [Medline: 33755661]
54. Parker C. Collection and rearing of container mosquitoes and a 24-h addition to the CDC bottle bioassay. J Insect Sci 2020;20(6):13 [FREE Full text] [doi: 10.1093/jisesa/ieaa059] [Medline: 33135763]
55. Lopez-Monroy B, Gutierrez-Rodriguez SM, Villanueva-Segura OK, Ponce-Garcia G, Morales-Forcada F, Alvarez LC, et al. Frequency and intensity of pyrethroid resistance through the CDC bottle bioassay and their association with the frequency of kdr mutations in aedes aegypti (Diptera: Culicidae) from mexico. Pest Manag Sci 2018;74(9):2176-2184. [doi: 10.1002/ps.4916] [Medline: 29575404]
56. Team W. Manual for monitoring insecticide resistance in mosquito vectors and selecting appropriate interventions. Control of Neglected Tropical Diseases (NTD) GMP (GMP). 2022. URL: https://tinyurl.com/yc8nbbw8 [accessed 2022-06-22]
57. Team W. Determining discriminating concentrations of insecticides for monitoring resistance in mosquitoes: report of a multi-centre laboratory study and WHO expert consultations. Control of Neglected Tropical Diseases (NTD) GMP (GMP). URL: https://tinyurl.com/yc8nbbw8 [accessed 2022-03-27]
58. Matthews BJ, Dudchenko O, Kingan SB, Koren S, Antoshechkin I, Crawford JE, et al. Improved reference genome of aedes aegypti informs arbovirus vector control. Nature 2018;563(7732):501-507 [FREE Full text] [doi: 10.1038/s41586-018-0692-z] [Medline: 30429615]
59. Badolo A, Sombié A, Yaméogo F, Wangrawa DW, Sanon A, Pignatelli PM, et al. First comprehensive analysis of aedes aegypti bionomics during an arbovirus outbreak in west africa: dengue in ouagadougou, burkina faso, 2016-2017. PLoS Negl Trop Dis 2022;16(7):e0010059 [FREE Full text] [doi: 10.1371/journal.pntd.0010059] [Medline: 35793379]
60. Souza-Neto JA, Powell JR, Bonizzoni M. Aedes aegypti vector competence studies: a review. Infect Genet Evol 2019;67:191-209 [FREE Full text] [doi: 10.1016/j.meegid.2018.11.009] [Medline: 30465912]
61. Angleró-Rodríguez YI, MacLeod HJ, Kang S, Carlson JS, Jupatanakul N, Dimopoulos G. Molecular responses to zika Virus: modulation of infection by the toll and Jak/Stat immune pathways and virus host factors. Front Microbiol 2017;8:2050 [FREE Full text] [doi: 10.3389/fmicb.2017.02050] [Medline: 29109710]
62. Dengue and severe dengue factsheet. World Health Organization (WHO). URL: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue [accessed 2024-04-23]
63. Dengue: guidelines for diagnosis, treatment, prevention and control. Organization. WH. World Health Organization. URL: https://www.who.int/neglected_diseases/resources/9789241547871/en/ [accessed 2024-04-02]
64. Zika virus. Organization WH. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/zika-virus [accessed 2024-04-04]

XSL•FO
RenderX

65. Wilder-Smith A, Gubler DJ, Weaver SC, Monath TP, Heymann DL, Scott TW. Epidemic arboviral diseases: priorities for research and public health. Lancet Infect Dis 2017;17(3):e101-e106. [doi: 10.1016/S1473-3099(16)30518-7] [Medline: 28011234]
66. Facchinelli L, Badolo A, McCall PJ. Biology and behaviour of in the human environment: opportunities for vector control of arbovirus transmission. Viruses 2023;15(3):636 [FREE Full text] [doi: 10.3390/v15030636] [Medline: 36992346]
67. Ouédraogo WM, Toé KH, Sombié A, Viana M, Bougouma C, Sanon A, et al. Impact of physicochemical parameters of aedes aegypti breeding habitats on mosquito productivity and the size of emerged adult mosquitoes in ouagadougou City, burkina faso. Parasit Vectors 2022;15(1):478 [FREE Full text] [doi: 10.1186/s13071-022-05558-3] [Medline: 36539816]
68. Tedjou AN, Kamgang B, Yougang AP, Wilson-Bahun TA, Njiokou F, Wondji CS. Patterns of ecological adaptation of and and Indices highlight the potential risk of arbovirus transmission in yaoundé, the capital City of cameroon. Pathogens 2020;9(6):491 [FREE Full text] [doi: 10.3390/pathogens9060491] [Medline: 32575721]
69. Badolo A, Ilboudo-Sanogo E, Ouédraogo AP, Costantini C. Evaluation of the sensitivity of aedes aegypti and anopheles gambiae complex mosquitoes to two insect repellents: DEET and KBR 3023. Trop Med Int Health 2004;9(3):330-334 [FREE Full text] [doi: 10.1111/j.1365-3156.2004.01206.x] [Medline: 14996361]
70. Rose NH, Badolo A, Sylla M, Akorli J, Otoo S, Gloria-Soria A, et al. Dating the origin and spread of specialization on human hosts in mosquitoes. Elife 2023;12:e83524 [FREE Full text] [doi: 10.7554/eLife.83524] [Medline: 36897062]
71. Gingras Y, Larivière V, Macaluso B, Robitaille JP. The effects of aging on researchers' publication and citation patterns. PLoS One 2008;3(12):e4048 [FREE Full text] [doi: 10.1371/journal.pone.0004048] [Medline: 19112502]
72. Gubler DJ. Dengue and dengue hemorrhagic fever. Clin Microbiol Rev 1998;11(3):480-496 [FREE Full text] [doi: 10.1128/CMR.11.3.480] [Medline: 9665979]
73. Bernatović I, Slavec Gomezel A, Černe M. Mapping the knowledge-hiding field and its future prospects: a bibliometric co-citation, co-word, and coupling analysis. Knowledge Management Research & Practice 2021;20(3):394-409. [doi: 10.1080/14778238.2021.1945963]
74. Zupic I, Čater T. Bibliometric methods in management and organization. Organizational Research Methods 2014;18(3):429-472. [doi: 10.1177/1094428114562629]
75. Small H. Co‐citation in the scientific literature: a new measure of the relationship between two documents. J. Am. Soc. Inf. Sci 2007;24(4):265-269. [doi: 10.1002/asi.4630240406]
76. Yan E, Wu C, Song M. The funding factor: a cross-disciplinary examination of the association between research funding and citation impact. Scientometrics 2017;115(1):369-384. [doi: 10.1007/s11192-017-2583-8]
77. Mota FB, Fonseca BDPFE, Galina AC, Silva RMD. Mapping the dengue scientific landscape worldwide: a bibliometric and network analysis. Mem Inst Oswaldo Cruz 2017;112(5):354-363 [FREE Full text] [doi: 10.1590/0074-02760160423] [Medline: 28443981]
78. Mali F, Pustovrh T, Platinovšek R, Kronegger L, Ferligoj A. The effects of funding and co-authorship on research performance in a small scientific community. Science and Public Policy 2016;44(4):486-496. [doi: 10.1093/scipol/scw076]
79. Hirv T. The interplay of the size of the research system, ways of collaboration, level, and method of funding in determining bibliometric outputs. Scientometrics 2022;127(3):1295-1316. [doi: 10.1007/s11192-021-04232-2]
80. Guo YH, Lai SJ, Liu XB, Li GC, Yu HJ, Liu QY. Governmental supervision and rapid detection on dengue vectors: an important role for dengue control in China. Acta Trop 2016;156:17-21. [doi: 10.1016/j.actatropica.2015.12.011] [Medline: 26739658]
81. Pan J, Fang C, Yan J, Yan H, Zhan B, Sun Y, et al. Chikungunya fever outbreak, zhejiang province, China, 2017. Emerg Infect Dis 2019;25(8):1589-1591 [FREE Full text] [doi: 10.3201/eid2508.181212] [Medline: 31310205]
82. Qiu J. China?s funding system and research innovation. Natl Sci Rev 2014;1(1):161-163.
83. Soosaraei M, Khasseh AA, Fakhar M, Hezarjaribi HZ. A decade bibliometric analysis of global research on leishmaniasis in web of science database. Ann Med Surg (Lond) 2018;26:30-37 [FREE Full text] [doi: 10.1016/j.amsu.2017.12.014] [Medline: 29387384]
84. Fontecha G, Sánchez A, Ortiz B. Publication trends in neglected tropical diseases of latin America and the caribbean: a bibliometric analysis. Pathogens 2021;10(3):356 [FREE Full text] [doi: 10.3390/pathogens10030356] [Medline: 33802834]
85. Zheng S, He A, Yu Y, Jiang L, Liang J, Wang P. Research trends and hotspots of health-related quality of life: a bibliometric analysis from 2000 to 2019. Health Qual Life Outcomes 2021;19(1):130 [FREE Full text] [doi: 10.1186/s12955-021-01767-z] [Medline: 33892744]

Abbreviations

AJTMH: American Journal of Tropical Medicine and Hygiene
AMCA: American Mosquito Control Association
JME: Journal of Medical Entomology
NTD: Neglected Tropical Diseases
RGR: relative growth rate
WHO: World Health Organization

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 27.08.24; peer-reviewed by DW Wangrawa, N Kumar, YP Sharma; comments to author 27.12.24; revised version received 26.01.25; accepted 12.02.25; published 23.04.25.

Please cite as:
Cheong YL, Mohd Ghazali S, Mat Hashim MH, Che Ibrahim MK, Amran A, Tiunh TY, Lim HL, Cheah YK, Gill BS, Lim KH
Exploring 97 Years of Aedes aegypti as the Vector for Dengue, Yellow Fever, Zika, and Chikungunya (Diptera: Culicidae): Scientometric Analysis
Interact J Med Res 2025;14:e65844
URL: https://www.i-jmr.org/2025/1/e65844
doi:10.2196/65844
PMID:

©Yoon Ling Cheong, Sumarni Mohd Ghazali, Mohd Hazilas Mat Hashim, Mohd Khairuddin Che Ibrahim, Afzufira Amran, Tsye Yih Tiunh, Hui Li Lim, Yong Kang Cheah, Balvinder Singh Gill, Kuang Hock Lim. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 23.04.2025. 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

Characterization of Global Research Trends and Prospects on Prone Positioning in Respiratory Failure: Bibliometric Analysis

Rong Lei*, MM; Feng Yue*, MM; Chaofu Yue*, MM; Zihan Zhang, MM; Xian Huang, MM; Qiaolin Li, MM; Zhigang Yang, MM; Rong Li, MM; Keyi Zhao, MM; Mei Yang, MM

The Qujing NO.1 People's Hospital, No. 1 Yuanlin Road, Qilin District, Qujing City, China

*these authors contributed equally

Corresponding Author:
Mei Yang, MM
The Qujing NO.1 People's Hospital, No. 1 Yuanlin Road, Qilin District, Qujing City, China

Abstract

Background: Prone positioning has emerged as a crucial intervention in managing acute respiratory failure, especially in acute respiratory distress syndrome and patients with COVID-19. Given the increasing interest in this field, it is important to characterize global research trends and key contributors to identify future research directions.

Objective: This study aimed to analyze global research trends, collaboration networks, and research hotspots related to prone positioning in respiratory failure through a comprehensive bibliometric analysis.

Methods: Bibliometric analyses were conducted using CiteSpace and Biblioshiny software on publications up to December 31, 2023, from the Web of Science Core Collection, focusing on prone positioning in respiratory failure.

Results: A total of 1263 research articles were identified, published in 50 countries by numerous institutions. The United States, France, and Germany contributed the most publications, with the United States producing 21.9% (275/1263) of the total. Key authors such as Claude Guerin and Luciano Gattinoni were identified as major contributors to the field. Keyword co-occurrence analysis revealed the dynamic nature of prone positioning research in respiratory failure. It highlighted protective ventilation and COVID-19-related acute respiratory distress syndrome as emerging hotspots, indicating a shift in focus during the pandemic.

Conclusions: This study revealed a rapidly growing body of literature on prone positioning in respiratory failure, especially in the context of COVID-19. The findings underscore the importance of further multicenter clinical trials to validate current practices and refine treatment protocols. In addition, the application of prone positioning in non-intubated patients represents a potential future research direction.

(Interact J Med Res 2025;14:e67276) doi:10.2196/67276

KEYWORDS

acute respiratory distress syndrome; respiratory failure; acute respiratory failure; ARF; bibliometric analysis; critical care; bibliometrics; lungs; prone positioning; COVID-19; SARS-CoV-2; coronavirus; infectious; pulmonary; pandemic

Introduction

Prone positioning, defined as a posture in which the patient lies face down with the chest and abdomen in contact with the surface and the back facing upwards, is commonly used in clinical settings for respiratory support and surgical procedures. Recently, prone positioning has gained significant attention as a therapeutic intervention, particularly in the management of acute respiratory failure (ARF) [1]. Its effectiveness is rooted in the interplay between body positioning and pulmonary mechanics, a relationship that has been explored extensively in recent decades. Historically, respiratory management in patients with ARF has focused on methods to enhance oxygenation and reduce complications [2].

In 1974, Mellins et al [3] investigated how physical therapies and body positioning, including the prone positioning, could influence lung volumes and facilitate better secretion clearance in pediatric patients with pulmonary diseases, supporting the notion that posture significantly impacts respiratory mechanics. Subsequently Bryan [4] offered a critical perspective on the mechanical implications of body positioning, and suggested that the supine position could exacerbate loss of lung volume in dependent regions due to gravitational effects on the diaphragm, leading to diminished ventilation and potential airway closure. Over the following decades, the role of prone positioning in improving oxygenation in patients with acute ARF was further investigated. In 2001, Gattinoni et al [5] conducted a randomized clinical trial on prone positioning in patients with acute respiratory distress syndrome (ARDS). While the study showed improvement in oxygenation, it did not demonstrate a significant survival benefit; however, it laid the groundwork for further investigation. The pivotal PROSEVA trial in 2013, conducted by Guérin et al [6], demonstrated that

XSL•FO
RenderX

early and prolonged prone positioning significantly reduced 28-day and 90-day mortality in severe patients with ARDS, leading to its recommendation as a standard practice for managing ARDS.

The COVID-19 pandemic has renewed attention to prone positioning, not only for intubated patients but also, notably, for awake, nonintubated patients experiencing hypoxemia [7]. Studies during the pandemic have confirmed that prone positioning can improve oxygenation in patients with COVID-19 and can reduce the need for endotracheal intubation, which is confirmed by computed tomography scan and electrical impedance tomography [8]. The cumulative findings from these studies have shaped current respiratory care practices, where prone positioning is now routinely used in ARF management. This evolution in understanding highlights the interplay between pulmonary mechanics, perfusion distribution, and body positioning, establishing the prone intervention as a valuable tool in clinical practice [9]. As the evidence supporting prone continues to grow, it becomes increasingly important to assess its broader impact on clinical practice, thereby guiding future research and refining its application across diverse clinical settings.

Bibliometric analysis provides a quantitative approach to assessing the development and dissemination of knowledge within a particular field [10]. By mapping the trajectory of research on prone positioning and respiratory failure, it is possible to identify key trends, influential studies, and gaps in the existing literature. This, in turn, offers valuable insights into the evolution of clinical practice and the global impact of research efforts [11]. Previous bibliometric studies in respiratory failure have addressed various topics, none have specifically focused on prone positioning in respiratory failure [12,13]. This study aims to fill this gap by performing a comprehensive bibliometric analysis of the existing literature on prone positioning and its application in respiratory failure. Such insights are essential for guiding future research directions, promoting international collaboration, and ensuring that the benefits of prone positioning are fully realized in diverse clinical settings.

Methods

Data Source and Search Strategy

The data for this bibliometric analysis were sourced from the Web of Science Core Collection (WoSCC). A comprehensive search was conducted on August 21, 2024, using the following search strategy: “prone positioning” AND (“acute respiratory failure” OR “ARDS” OR “respiratory failure” OR “acute respiratory distress syndrome”). This strategy was designed capture studies on prone positioning in respiratory failure, using both general and specific terms. The initial search identified a total of 2143 records.

Inclusion and Exclusion Criteria

To refine the dataset for bibliometric analysis, specific inclusion and exclusion criteria were applied. Initially, 108 records from the year 2024 were excluded to ensure the analysis focused on more established literature, resulting in 2035 articles. The dataset was further refined by excluding the following types of publications: 360 review articles, 126 meeting abstracts, 121 editorial materials, 114 letters, 43 proceeding papers, 5 corrections, and 3 early access articles. These exclusions were made to retain only original research articles, which are more likely to contribute directly to the understanding and development of prone positioning in respiratory failure. After applying these criteria, 1263 articles were selected for subsequent analysis, and no duplicates were found.

Data Cleaning and Preprocessing

To ensure the accuracy and consistency of our dataset, we conducted a comprehensive data cleaning and preprocessing procedure. This process involved the removal of duplicate records, identified through digital object identifiers and study titles. We used the “biblioshiny” package in RStudio to standardize author and institutional names, consolidating any variations. In addition, the merge function was used to address synonyms, aliases, and singularor plural discrepancies in keywords. Noninformative keywords, such as “care,” “failure,” and “trial,” were removed using the remove function.

Bibliometric Analysis

Our bibliometric analysis followed a broad-to-specific framework, including an overview, countries or districts, authors or institutions, journals or most cited publications, keywords and trends. The collected data were imported into Biblioshiny (R version 4.3.3; Institute for Statistics and Mathematics, Vienna, Austria) [14], CiteSpace (6.3.R1; Drexel University, Philadelphia, PA, the United States for social network analysis and Burst detection [15], and Microsoft Office Excel 2016 for further analysis.

Biblioshiny is primarily used to visualize the analysis of global publication trends, countries, sources, and influential documents [16]. A range of bibliometric indicators are analyzed through biblioshiny to assess the output of countries, and journals. The number of publications was used to quantify productivity, while TC reflected academic influence. Local citations were examined to assess impact within specific subject areas. To evaluate the impact of sources, we assessed theh-index, g-index, and m-index [17]. The h-index provides a balanced view of productivity and citation impact, while the g-index emphasizes highly-cited articles. The m-index accounts for publication longevity, normalizing for years of activity. CiteSpace is a Java-based software developed by Chen for bibliometric analysis, supporting knowledge mining and data visualization. We used it to conduct author cocitation analysis, institution cocitation analysis, and keyword co-occurrence analysis. Visual network graphs were used to illustrate the relationships among nodes, where larger node sizes indicate a greater number of publications within a particular research area [18]. We also conducted keyword burst analysis to identify emerging research themes over time [19].

Results

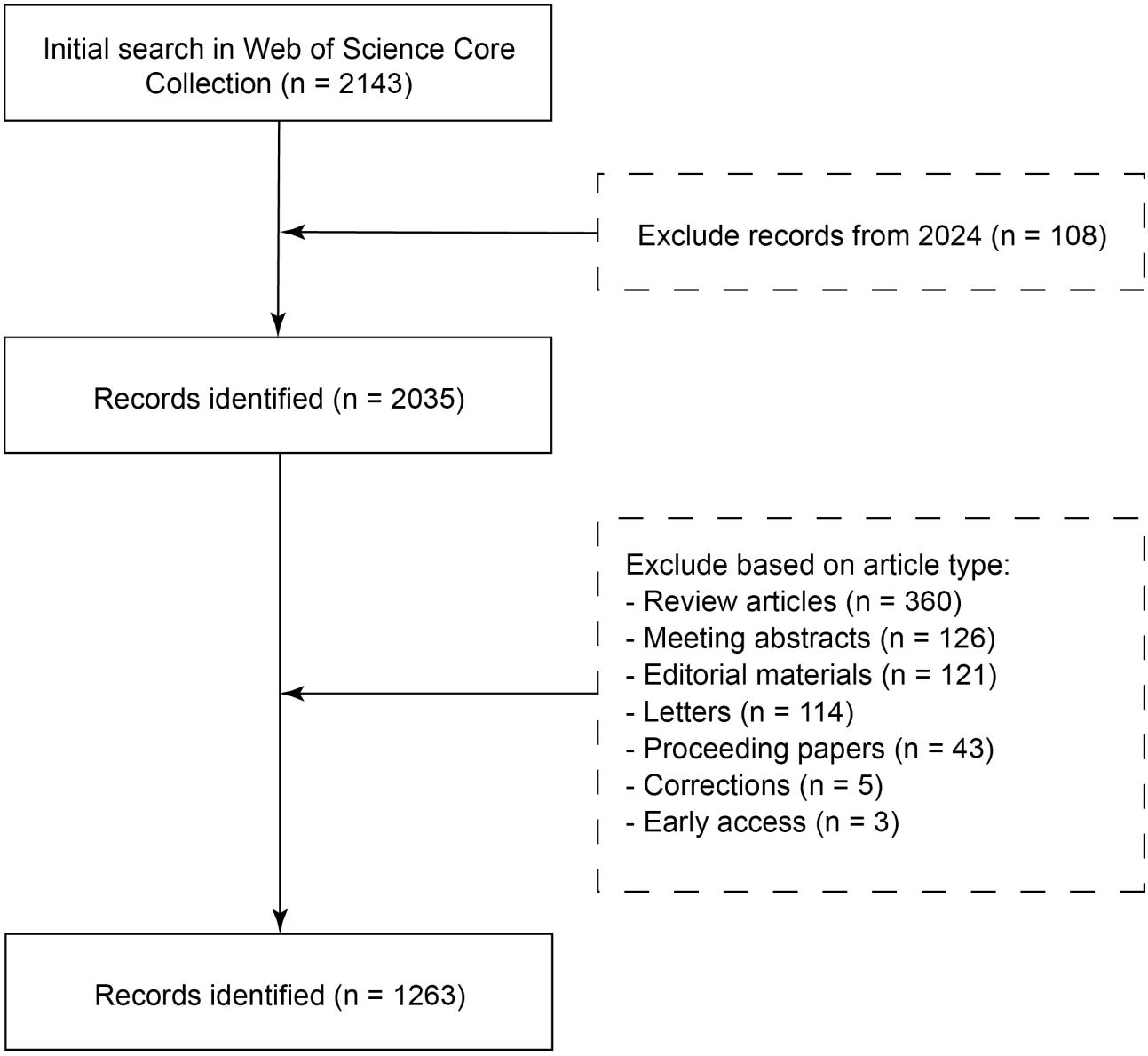
Search Results

A total of 2143 publications were identified from WoSCC. After excluding records from 2024 and filtering based on article type.

XSL•FO
RenderX

Finally, 1263 studies were identified for bibliometric analysis (Figure 1).

Figure 1. Record identification and selection.

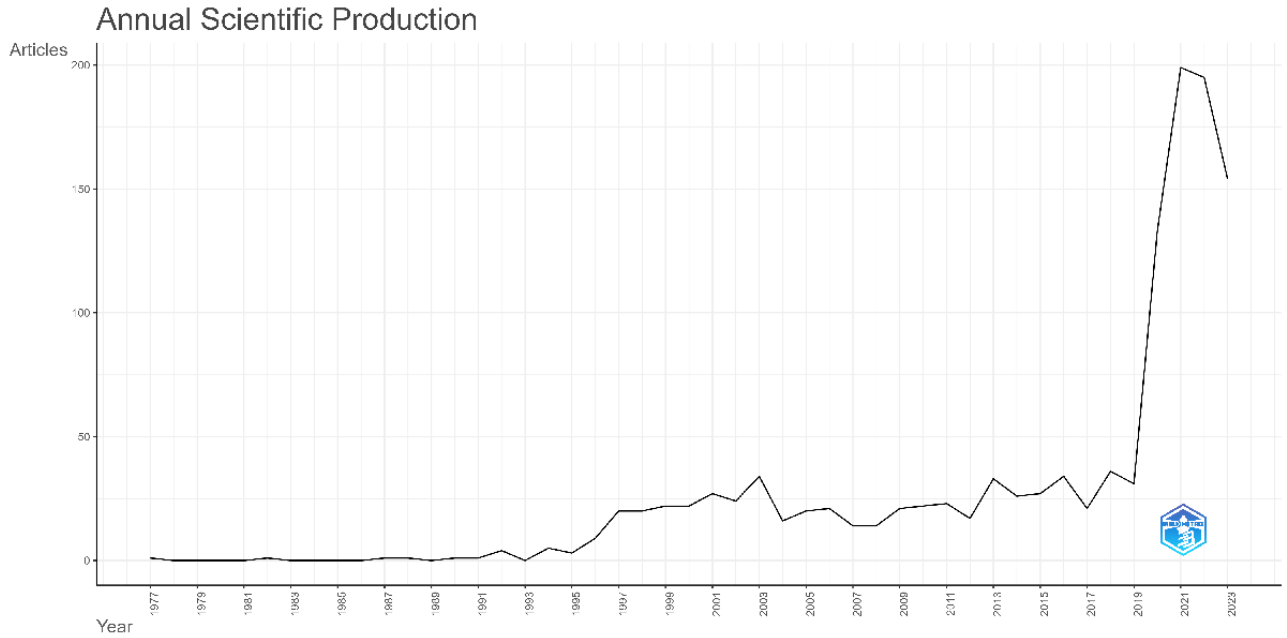


Global Publication Trends

The first publication on prone positioning in respiratory failure appeared in 1977, reporting that prone positioning significantly improves arterial oxygenation, allowing for reduced inspired oxygen concentrations in patients with ARF. In addition, there was only 1 to 5 publications per year before 1995. From 1995 to 2018, the number of publications showed a slow upward trend, whereas between2019 and 2023, there was a sharp increase, beginning with 31 publications in 2019 and peaking at 199 publications in 2021 (Figure 2).

XSL•FO
RenderX

Figure 2. Annual scientific production. This figure shows the annual number of publications on prone positioning in respiratory failure from 1977 to 2023. The first publication appeared in 1977, with only 1 to 5 articles per year before 1995. From 1995 to 2018, there was a gradual increase in publications. A significant rise in the number of publications occurred from 2019 to 2023, with 199 publications in 2021.



Countries and Districts

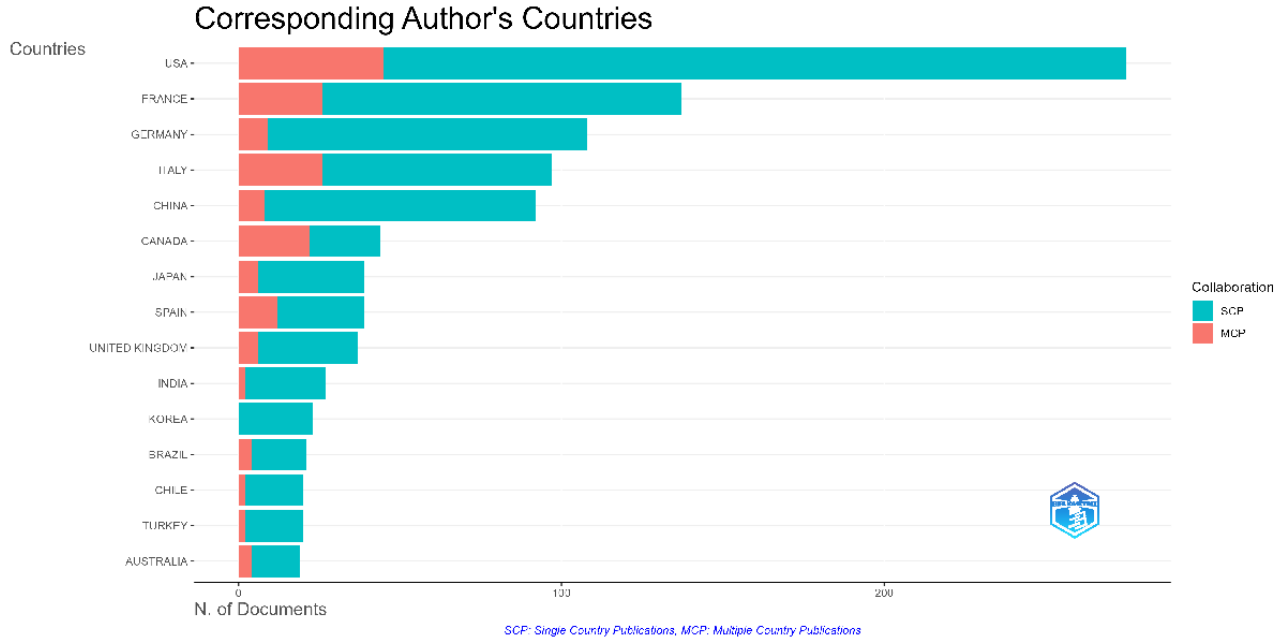
A total of 57 countries contributed to publications on prone positioning in respiratory failure. The United States was the leading contributor, with 275 publications (275/1253, 21.9% of the total), followed by France with 137 publications (137/1253, 10.9%), and Germany with 108 publications (108/1253, 8.6%). Italy contributed 97 publications (97/1253, 7.7%), and China contributed 92 publications (92/1253, 7.3%). Other major contributors included Canada (44/1253; 3.5%), Japan (39/1253; 3.1%), Spain (39/1253; 3.1%), the United Kingdom (37/1253; 3.0%), and India (27/1253; 2.2%).

In terms of citation impact, the United States had the highest total number of citations (10,146) with an average of 36.9 citations per document. France followed with 9229 citations and an average of 67.4 citations per document. Notably, Canada, despite contributing only 3.5% (44/1253) of the publications, had the highest average citations per document, reaching 132.5.

The analysis of corresponding authors' countries revealed varying patterns of domestic and international collaboration (Figure 3; Sheet 1 in Multimedia Appendix 1). The United States had the highest total number of publications, with 230 single-country publications and 45 multiple-country publications (MCPs), resulting in an MCP rate of 16.4% (45/275). Although Canada had a relatively lower publication count (44 articles), it exhibited the highest MCP rate among the top contributing countries, with 50% of its publications involving international collaboration. By contrast, India had the lowest MCP rate at 7.4% (2/27), reflecting a higher proportion of domestically authored publications.

XSL•FO
RenderX

Figure 3. Corresponding author's country distribution. This figure presents the distribution of single-country publications and multiple-country publications by country.

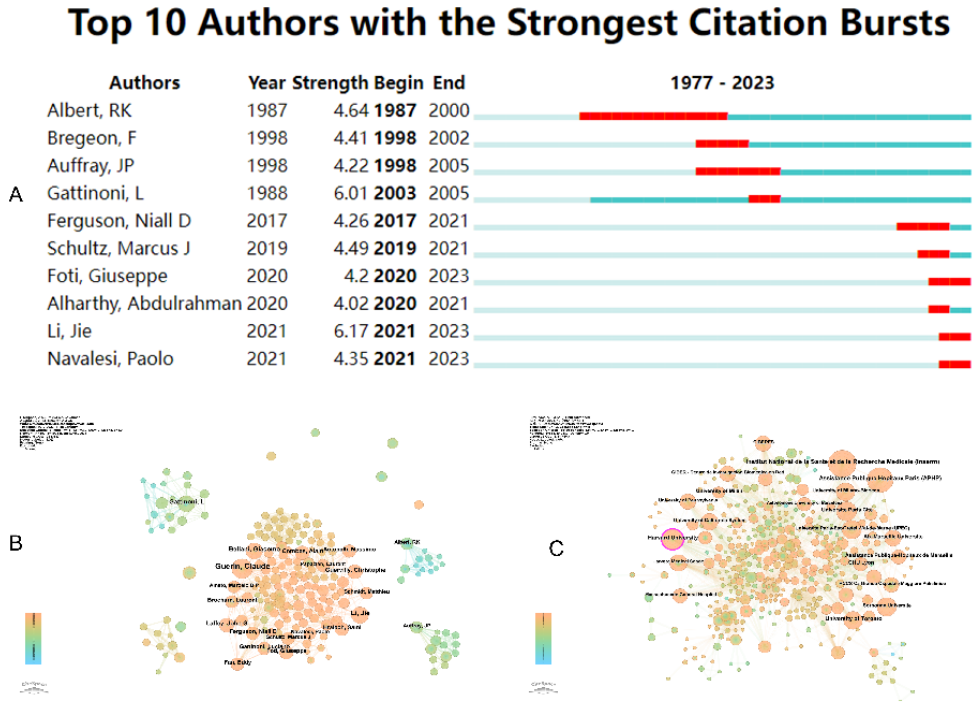


Authors and Institutions

A total of 553 authors contributed to publications on prone positioning and respiratory failure. Albert et al [20-23] published 6 articles and experienced a notable citation burst beginning in 1987 and continuing until 2000, with a burst strength of 4.64 (Figure 4A). Papazian et al [24,23] also experienced a citation burst from 1998 to 2005, with a strength of 4.22. Gattinoni demonstrated the highest citation burst strength of 6.01, occurring between 2003 and 2005. With 10 publications, Gattinoni has been a leading figure in this field since 1988, particularly in research on ARDS and prone positioning. More recently, Ehrmann et al [25] and Cammarota [26] have shown emerging influence, with citation bursts beginning in 2021 and continuing through 2023, and burst strengths of 6.17 and 4.35, respectively. Network analysis identified Claude Guérin as the most prolific and central author, with 16 publications and a centrality score of 0.03, underscoring his influential role in this research domain (Sheet 2 in Multimedia Appendix 1). Guérin's contributions have been especially significant since 2006 (Figure 4B). In the early stages of the field, authors such as Gattinoni et al [27], and Michelet et al [23] conducted research in relatively isolated clusters. However, since the onset of the COVID-19 pandemic, collaboration among researchers has increased notably, reflecting a shift toward more integrated global research.networks. Institut National de la Santé et de la Recherche Médicale and Assistance Publique–Hôpitaux de Paris, with 82 and 70 publications respectively, rank among the top institutions contributing to research on prone positioning and respiratory failure (Figure 4C). Harvard University leads with 51 publications and exhibits the highest centrality score (0.15), indicating its pivotal role in connecting diverse research clusters within the global collaboration network. Other institutions, such as Leipzig University and the University of Milan, also demonstrate relatively high centrality values (0.09 and 0.08, respectively), suggesting that they serve as important bridges across international research communities (Sheets 3 in Multimedia Appendix 1). An emerging contributor, Southeast University in China, reflects the growing involvement of Chinese institutions in this research domain.The analysis further reveals distinct collaboration patterns, with some institutions predominantly engaging in domestic (intracountry) collaborations, while others—such as Harvard University—participate extensively in international research partnerships, advancing global progress in the study of prone positioning.

XSL•FO
RenderX

Figure 4. Authors and institutions. (**A**) Top 10 Authors with the Strongest Citation Bursts. This panel displays the top 10 authors with the strongest citation bursts in the field of prone positioning and respiratory failure. The red line segments represent the time periods of citation bursts, while the burst strength indicates the intensity of citation activity, with higher values reflecting stronger bursts. (**B**) Co-authorship network Visualization. This panel visualizes the co-authorship network in prone positioning research. Nodes represent authors, and edges between nodes represent co-authorship relationships. The network highlights key authors and their collaboration patterns over time.(**C**) Network of Co-authors' Institution Visualization. This panel illustrates the co-authorship network at the institutional level. Nodes represent institutions, and edges between them reflect collaborative publishing efforts. The network highlights prominent institutions and their involvement in global research initiatives on prone positioning and respiratory failure.

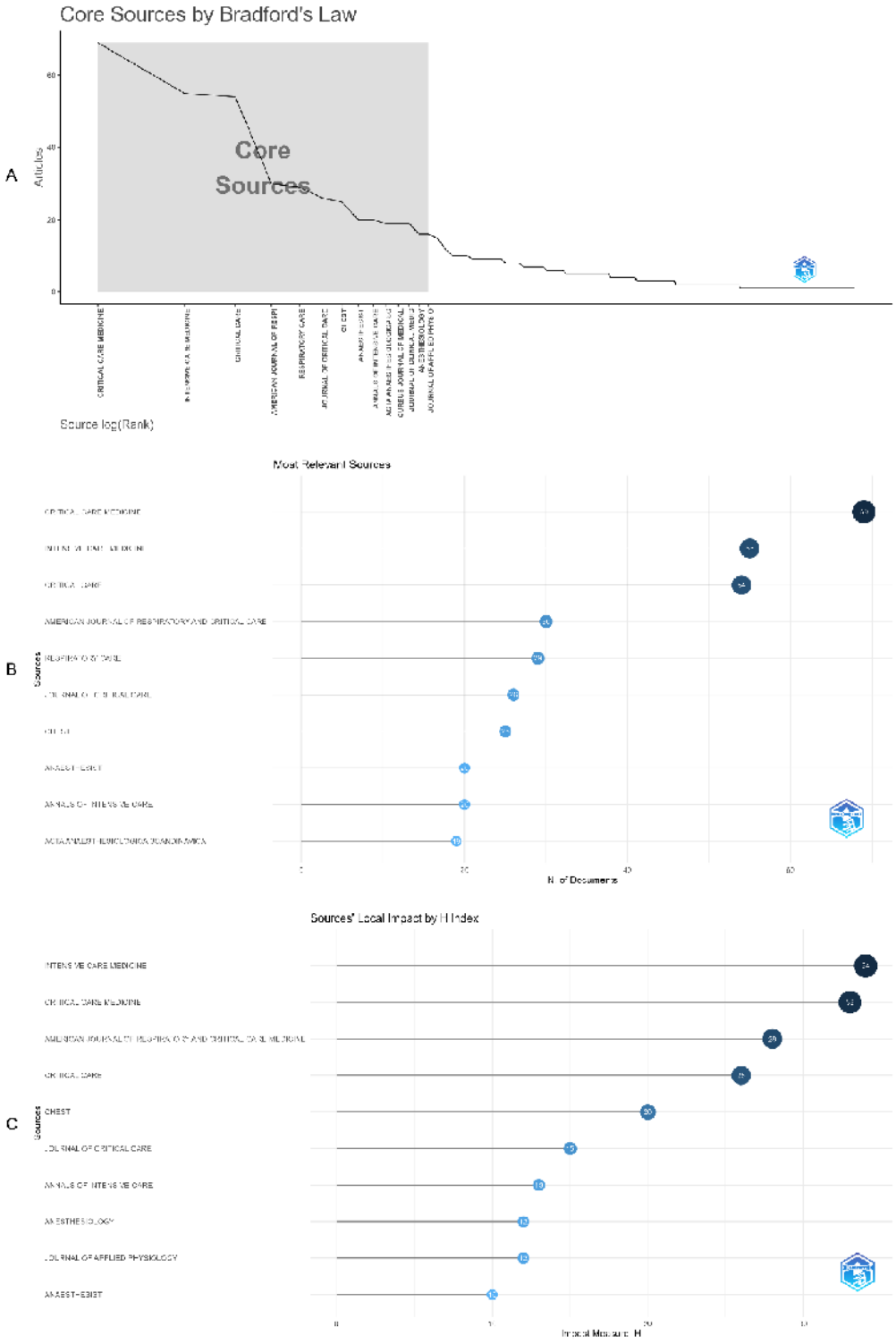


Journals and Most Cited Publications

Bradford's Law provides a framework for quantifying the distribution of articles across journals within a specific research field. It posits that publications can be grouped into a core set of journals contributing the largest number of articles, followed by successive zones with decreasing productivity. According to Bradford's Law Analysis, 13 journals were identified as core journals, The leading journal, *Critical Care Medicine*, has the highest frequency with 69 articles, followed by *Intensive Care Medicine* (55 articles), and *Critical Care* (54 articles). These journals constitute the core outlets for high-impact and frequently cited research on prone positioning and respiratory failure (Figure 5A). Further analysis of relevant sources confirmed *Critical Care Medicin*e as the most productive journal, followed closely by *Intensive Care Medicine* and *Critical Care*. *The American Journal of Respiratory and Critical Care Medicine* also ranked highly, with 30 articles published in this area (Figure 5B). In terms of source impact, *Intensive Care Medicine* demonstrated the highest h-index of 34, along with the highest TC of 5925. This was closely followed by *Critical Care Medicine* (h-index 33, TC 4307) and *The American Journal of Respiratory and Critical Care Medicine* (h-index 28, TC 4726) (Figure 5C).

XSL•FO
RenderX

Figure 5. Authors and institutions. (**A**) Bradford Law bibliometrix. (**B**) Most relevant journals. (**C**) Impact index of journals.



High values across h-index, g-index, and m-index indicate that these journals exhibit both substantial productivity and significant impact in the field of respiratory failure, particularly concerning prone positioning (Sheet 3 in Multimedia Appendix 1). *Critical Care Medicine* and *Intensive Care Medicine* had comparable m-index values of 1.179 and 1.172, respectively, suggesting sustained academic influence since their emergence in the mid-to-late 1990s. Their consistent growth in citations and publication output reflects their central role in disseminating research in this area (Sheet 4 in Multimedia Appendix 1). Despite its relatively recent entry into the field (established in 2015), *Annals of Intensive Care* recorded the highest m-index (1.3), indicating rapid growth in influence. In contrast, older, more established journals such as *CHEST* and *The Journal of Applied Physiology* have shown steadier but slower growth in impact over time.

The most cited publication in this analysis is the study by Bellani et al [28], published in *JAMA*, which received 3240 TCs, with

XSL•FO
RenderX

an average of 360 citations per year (TC/Y) and a normalized TC score of 23.25 (Table 1). This pivotal study presents large-scale data from the LUNG SAFE study, offering critical insights into the epidemiology, management, and outcomes of ARDS. Closely following is the work by Guérin et al (2013), published in *The New England Journal of Medicine*, with 2423 TC, a TC/Y of 201.92, and a normalized TC of 12.85. This landmark clinical trial established the efficacy of prone positioning in improving survival rates among patients with severe ARDS, and it fundamentally altered international treatment guidelines. A more recent publication by Schmidt et al (2021), appearing in *Intensive Care Medicine*, has already accumulated 476 citations, with a remarkable TC/Y of 119.00 and a normalized TC of 22.80. The rapid accumulation of citations reflects the high relevance of this study, particularly in the context of the COVID-19 pandemic, during which the management of respiratory failure and ARDS has become a global research priority.

Table . The most cited publication by top 10-record.

Rank	Title	TC[a]	TC per year	Normalized TC
1	Epidemiology, patterns of care, and mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries	3240	360.00	23.25
2	Prone positioning in severe acute respiratory distress syndrome	2423	201.92	12.85
3	Efficacy and economic assessment of conventional ventilatory support versus extracorporeal membrane oxygenation for severe adult respiratory failure (CESAR): a multicentre randomised controlled trial	2301	143.81	15.63
4	Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock, 2012	1426	118.83	7.56
5	Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012	933	77.75	4.95
6	An official american thoracic society/European Society of intensive care medicine/society of critical care medicine clinical practice guideline: mechanical ventilation in adult patients with acute respiratory distress syndrome	929	116.13	13.84
7	Effect of prone positioning on the survival of patients with acute respiratory failure	733	30.54	14.28
8	Extracorporeal membrane oxygenation for ARDS in adults	575	41.07	12.34
9	The concept of “baby lung”	505	25.25	5.91
10	Clinical characteristics and day-90 outcomes of 4244 critically ill adults with COVID-19: a prospective cohort study	476	119.00	22.80

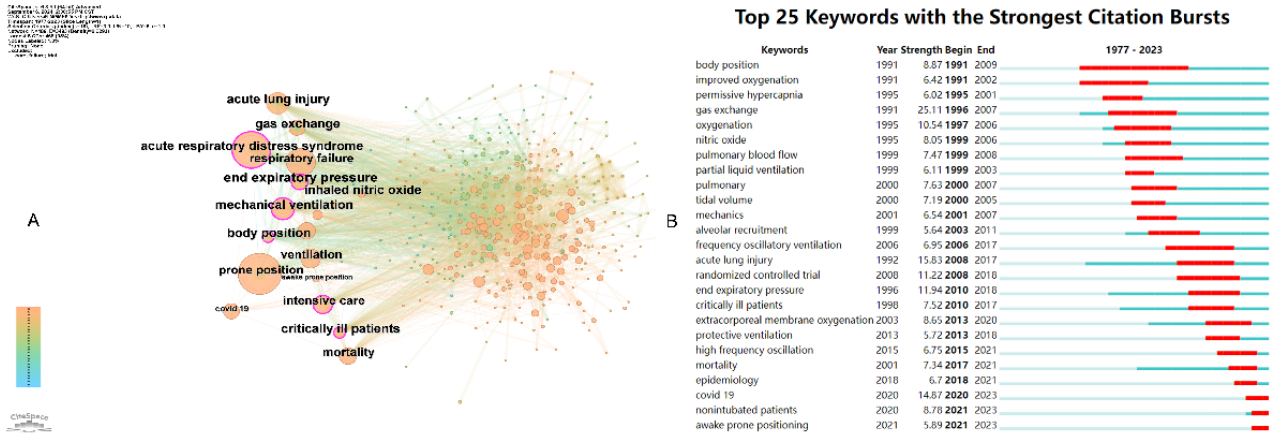
[a]TC: total citation.

XSL•FO
RenderX

Co-Occurring Author Keywords

In the analysis of co-occurring author keywords related to prone positioning and respiratory failure, several critical terms emerged as focal points, reflecting the research trends and clinical focus areas over the years. Prone position (count: 592 and centrality: 0.08), was the most frequently occurring keyword, consistently central to discussions on improving oxygenation and outcomes in patients with ARDS and respiratory failure (Figure 6A). ARDS (count: 475, centrality: 0.1) is the second most frequent keyword, highlighting the prominence of ARDS in the context of respiratory failure, particularly in mechanically ventilated patients (Sheet 5 in Multimedia Appendix 1). The co-occurrence of mechanical ventilation (count: 292 and centrality: 0.12) further emphasizes the critical relationship between mechanical support strategies and the management of ARDS and respiratory failure. The inclusion of extracorporeal membrane oxygenation (count: 111) and inhaled nitric oxide (count: 79) among the top keywords reflects the exploration of advanced therapies for refractory hypoxemia, often in conjunction with prone positioning. Notably, COVID-19 (count: 58) and awake prone positioning (count: 18) have emerged as recent significant keywords, reflecting the shift in clinical focus due to the pandemic and the adaptation of prone positioning in nonintubated, awake patients to improve oxygenation without invasive mechanical ventilation.

Figure 6. Co-occurring author keywords analysis. (**A**) Co-occurring author keywords network Visualization. This panel visualizes the network of co-occurring author keywords related to prone positioning and respiratory failure. Nodes represent individual keywords, and edges between nodes indicate the co-occurrence relationships. The network highlights the most frequently mentioned and connected terms in the research field.(**B**) Top 25 Keywords with the Strongest Citation Bursts. This panel presents the top 25 keywords with the strongest citation bursts, reflecting trends in the research over time. The red line segments represent the time periods of citation bursts, while the burst strength indicates the intensity of citation activity, with higher values reflecting stronger bursts.



Based on the keywords burst analysis, the top 25 keywords revealed clear trends in the evolution of research on prone positioning and respiratory failure (Figure 6B).

Early Stage (around 2000)

During this period, keywords such as body position and gas exchange were prominent. These terms suggest that the research focus was centered on understanding the pathophysiology of ARDS and improving treatment strategies—particularly by optimizing mechanical ventilation. Researchers aimed to enhance survival rates among patients with ARDS by refining ventilatory support techniques.

Middle Stage (around 2010)

In this phase, the emphasis shifted toward prone position and lung protective strategy, indicating a growing consensus on the physiological benefits of prone positioning and its role within lung-protective ventilation frameworks.

Recent Stage (around 2020)

In the most recent phase, keywords such as COVID-19, awake prone positioning, and nonintubated patients emerged strongly. These bursts reflect the global urgency in treating ARDS in the context of the COVID-19 pandemic. The renewed interest in prone positioning—particularly for awake, nonintubated patients—underscores its adoption as a noninvasive, life-saving intervention for COVID-19-related respiratory failure.

Discussion

Principal Findings

The bibliometric analysis of research on prone positioning in respiratory failure offers a comprehensive view of the evolution, global trends, key contributors, and future directions in this important clinical area. Prone positioning has evolved over the past 5 decades from a theoretical concept to a widely adopted therapeutic intervention [29,30]. Early observational studies demonstrated its potential to improve oxygenation, a finding later corroborated by clinical trials [2,3]. Subsequently, numerous cohort studies and randomized controlled trials, confirmed its efficacy in respiratory failure [31-34]. The landmark PROSEVA trial in 2013 demonstrated a significant reduction in 28-day and 90-day mortality rates when early and prolonged prone positioning was applied to patients with severe ARDS [6]. This study established prone positioning as a standard intervention in patients with severe ARDS, where conventional supine ventilation often fails to maintain adequate oxygenation. With the onset of the COVID-19 pandemic, prone positioning was rapidly adopted in hypoxemic respiratory failure related to COVID-19. Studies have shown that prone positioning can improve oxygenation not only in patients with mechanically

XSL•FO
RenderX

ventilation but also in nonintubated, awake patients, suggesting a role in potentially delaying or preventing the need for invasive mechanical ventilation [35]. This trend underscores the importance of prone positioning in both traditional ARDS management and in the novel context of COVID-19, driving a surge in research output.

National Contributions and Global Distribution

The United States leads global contributions to research on prone positioning in respiratory failure, accounting for 21.9% of the total publications (Table 2). The United States research output is not only substantial in volume but also impactful, as evidenced by its 10,146 citations and an average citation per document of 36.9. This aligns with the country's leadership in critical care and respiratory research, as highlighted by key studies such as the LUNG SAFE study by Bellani et al [28]. The prominent role of European countries, including Germany, Italy, and France, reflects strong collaborations within the EU and international research networks, as exemplified by the PROSEVA trial [6]. Overall, the research landscape in prone positioning and respiratory failure is characterized by strong contributions from a mix of Western and Asian countries, with the United States, France, and Italy leading in terms of both quantity and impact of publications. The high average citation rates in countries like France, Canada, and the United Kingdom suggest that these nations produce particularly influential research, while emerging contributions from China and India highlight the increasing global interest and participation in this critical area of study [25,36-38].

Table . Distribution by country.

Rank	Country	Documents, n (%)	Citations, n	Average citations
1	United States	275 (21.9)	10146	36.90
2	France	137 (10.9)	9229	67.40
3..	Germany	108 (8.6)	2232	20.70
4	Italy	97 (7.7)	6008	61.90
5	China	92 (7.3)	1039	11.30
6	Canada	44 (3.5)	5831	132.50
7	Japan	39 (3.1)	348	8.90
8	Spain	39 (3.1)	1529	39.20
9	United Kingdom	37 (3)	3829	103.50
10	India	27 (2.2)	221	8.20

Influential Authors and Scientific Impact

The analysis identified several key authors who have made foundational contributions to the field of prone positioning and respiratory failure. (Figure 4). In the early stage Albert et al have significantly advanced the understanding of the physiological effects and clinical benefits of prone positioning in managing ARF, particularly in ARDS. His work spans from animal models to human studies, elucidating mechanisms that improve gas exchange and reduce lung injury, and identifying that one of the key benefits of prone positioning was the alleviation of lung compression caused by the heart, which have shaped modern critical care practices [20-22,39]. The researcher from France, Auffray JP has played a significant role in advancing the understanding of prone positioning and its interaction with other treatments for ARDS. Auffray's work, particularly from the late 1990s to the early 2000s, focused on exploring how the prone position, combined with adjunctive therapies like inhaled nitric oxide and pulmonary vasoconstrictors, could enhance oxygenation and improve clinical outcomes in patients with ARDS [23,24,40]. And his studies explored the variability in patient response to prone positioning, indicated that a short-term trial of prone positioning may not be sufficient to predict which patients will respond to the prone positioning therapy, showing the complexity of patient variability in ARDS management. He also examined the tomographic features of patients with ARDS, while prone positioning improved oxygenation in many cases, computed tomography scans could not reliably predict responders based on lung opacity [41,42].

Gattinoni L's research [43] on ARDS spans over several decades, beginning in the late 1980s and continuing into the present. His pioneering work has evolved alongside advances in medical technology and clinical understanding of respiratory failure. His earliest influential study on prone positioning was published in 1988, where he investigated how body position could improve oxygenation in patients with ARDS [43]. Throughout the 1990s and 2000s, Gattinoni's research expanded to establish the concept of the "baby lung," which revolutionized the approach to mechanical ventilation by emphasizing the need to protect the small, aerated portion of the lungs in patients with ARDS. This work also laid the foundation for understanding ventilator-induced lung injury and its prevention through lung-protective strategies [44]. Gattinoni, L's research into prone positioning, lung mechanics, and the differentiation between pulmonary and extrapulmonary ARDS continued well into the 2000s, with several landmark studies published in 2003. These studies helped shape modern ventilatory practices, emphasizing the importance of individualized ventilation settings and protective strategies [45-47]. His body of work remains highly influential in guiding the clinical management of ARDS, particularly in the context of critical care and mechanical ventilation.

XSL•FO
RenderX

Dr Claude Guérin, with the most publications and significant centrality in the coauthorship network, stands out as a pivotal figure in this area, particularly due to his work on the landmark PROSEVA trial, which demonstrated the survival benefit of prone positioning in patients with severe ARDS, provides robust evidence that prone positioning not only improved oxygenation but also led to significant reductions in 28-day and 90-day mortality rates [6]. Following the PROSEVA trial, Guérin continued to explore the physiological mechanisms and clinical applications of prone positioning. His subsequent studies addressed various aspects of prone positioning, including its effects on hemodynamics and respiratory mechanics [48]. Throughout his career, Guérin's research has consistently focused on optimizing the use of prone positioning and understanding its physiological benefits. His work has shaped clinical guidelines and transformed the standard of care for patients with ARDS, particularly in the context of the COVID-19 pandemic. The breadth and depth of his research, ranging from clinical trials to mechanistic studies, underscore his pivotal role in advancing critical care practices for respiratory failure [49]. The citation burst of more recent authors like Li Jie and Navalesi Paolo indicates that research on prone positioning continues to evolve, with current studies exploring newer dimensions, such as its use in nonintubated patients during the COVID-19 pandemic [25,50].

Institutional Landscape and Collaborative Networks

From an institutional perspective, several significant trends emerge that illuminate the global research landscape and collaboration dynamics in the field of prone positioning and respiratory failure. First, it is evident that highly productive authors are often affiliated with prestigious academic or medical institutions that drive global research efforts. For instance, Claude Guérin, one of the most prominent authors in this field, is affiliated with University of Lyon in France, which have contributed substantially to the field through several influential studies [6,51,52]. Institut National de la Santé et de la Recherche Médicale and Assistance Publique–Hôpitaux de Paris have been leaders in research output, while institutions such as Harvard University, Leipzig University, and the University of Milan exhibit high centrality, indicating their pivotal role in connecting global research networks [5,28,53-56]. These institutions often serve as research hubs due to their access to extensive clinical trial data and resources for conducting high-impact studies

Moreover, international collaboration is a hallmark of research in this field. European institutions, particularly those in France and Italy, demonstrate strong collaborative networks with North American institutions, such as Harvard University and the University of Toronto. These collaborations facilitate the sharing of clinical insights and promote the development of comprehensive guidelines, as evidenced by the high citation rates of multinational studies such as those by Guérin et al [6,49]. Chinese institutions, such as Southeast University, have shown a rapid increase in publication volume, signaling an emerging influence in global respiratory research [57,58]. These institutions drive forward clinical guidelines and research agendas, supported by strong international collaborations that enhance the global understanding of prone positioning in respiratory failure.

Research Focus and Keyword Evolution

The evolution of research keywords provides valuable insight into the shifting focus and progression of scientific inquiry in this field [59]. By analyzing keyword trends over time, distinct phases of research development can be observed, revealing the core themes and emerging topics that have shaped the understanding of prone positioning as a therapeutic strategy in respiratory failure. Early studies primarily explored the effect of prone positioning on oxygenation in patients with ARDS, with mechanical ventilation being at the forefront of therapeutic interventions. During this period, research often emphasized the physiological benefits of prone positioning, particularly its ability to improve ventilation-perfusion matching and reduce shunt fraction [27,37,44,60,61]. In the middle phase, research shifted toward prone positioning and lung protective strategies. Prone positioning was increasingly recognized not only for its oxygenation benefits but also for its role in mitigating ventilator-induced lung injury [61-63]. The onset of the COVID-19 pandemic marked a new phase in the research on prone positioning, with keywords such as "COVID-19" and "awake prone positioning" from 2020 onwards. This period saw an explosion of research, as prone positioning was adapted for use in both intubated and nonintubated patients with COVID-19-induced respiratory failure [64-66]. Awake prone positioning, in particular, emerged as a critical intervention to delay or prevent intubation in hypoxemic patients with COVID-19. This technique was rapidly incorporated into clinical practice due to its simplicity, cost-effectiveness, and demonstrated efficacy in improving oxygenation without the need for mechanical ventilation [38,67].

Research Challenges and Difficulties

Research on different body positions in the management of respiratory failure has evolved significantly over the years, with various positions being explored for their effects on oxygenation and ventilation. The supine position, while commonly used, can result in poor lung expansion in the dorsal regions due to the effects of gravity, leading to compression and collapse of these areas [68]. This can exacerbate ventilation inefficiency, particularly in patients with ARDS, where the lungs are already compromised by inflammation and fluid accumulation [27]. In contrast, prone positioning reverses the gravitational effects on the lungs. This position redistributes lung inflation along the gravitational axis, opening the collapsed dorsal lung regions while compressing the ventral regions. As a result, prone positioning enhances ventilation in the dorsal parts, which are typically underventilated in the supine position, thereby improving oxygenation. Yi Xin et al [69] demonstrated in a large animal model that prone positioning improves lung reinflation and recruitment, particularly in the dorso-caudal regions of the lung. Subsequent clinical applications have confirmed the effectiveness of prone positioning, and multiple clinical trials have been conducted to investigate its safety and efficacy in diverse patient populations [29]. However, despite its proven benefits, the application of prone positioning still faces several challenges. One major difficulty is the potential for complications associated with the physical maneuvering of patients into the prone position, which can result in inadvertent extubation or displacement of endotracheal tubes [70]. In

XSL•FO
RenderX

addition, patients with certain comorbid conditions, such as obesity, cardiovascular instability, or spinal deformities, may not tolerate prone positioning as effectively. Obesity, in particular, complicates the effectiveness of prone positioning due to altered lung mechanics and reduced functional residual capacity, making it more difficult to achieve adequate lung recruitment during positioning [71]. Furthermore, prolonged prone positioning can increase the risk of pressure ulcers and facial injuries, especially in patients who cannot maintain the position without significant support [9]. These challenges underscore the need for careful patient selection, individualized treatment plans, and vigilant monitoring during the application of prone positioning, particularly in the presence of severe comorbidities. Recent studies suggest that alternative techniques, such as sequential lateral positioning and bed tilting, can also improve oxygenation and serve as viable substitutes for prone positioning [72]. Sequential lateral positioning involves alternating between supine and lateral positions to improve lung recruitment, redistribute ventilation, and enhance oxygenation in patients with ARDS. Adjusting the tilt of the bed also enhances alveolar recruitment and improves blood flow redistribution; however, tilting beyond 30° may cause overdistension and lead to hemodynamic instability [73].

Recommendations for Future Work

Future research should focus on several key areas to advance the understanding and application of prone positioning in respiratory failure.

Awake Prone Positioning

While the benefits of prone positioning in intubated patients with ARDS are well-documented, more research is needed to understand its efficacy in nonintubated patients. The COVID-19 pandemic has prompted the use of awake prone positioning, but studies have reported inconsistent or even contradictory findings, highlighting the need for robust, multinational clinical trials to validate the results.

Personalized Prone Positioning Protocols

Research should explore whether personalized approaches to prone positioning, tailored to individual patient characteristics (eg, lung morphology, disease severity, lifestyle habits, nutritional status, BMI), can improve outcomes. Developing personalized prone positioning protocols could lead to more effective and targeted treatments for diverse patient populations.

Explore the Mechanisms Behind Prone Position

Future research should focus on understanding the physiological mechanisms underlying the effects of prone positioning on respiratory mechanics and cardiopulmonary interactions in patients with respiratory failure. Emerging technologies, such as electrical impedance tomography, offer opportunities for real-time monitoring of lung recruitment and ventilation-perfusion matching during prone positioning, which can further elucidate the physiological impact of this intervention.

Long-Term Outcomes

Most research to date has focused on short-term outcomes, such as oxygenation and survival rates. Future studies should examine the long-term impact of prone positioning on pulmonary function, quality of life, and psychological well-being in survivors.

International Collaboration

Given the complex, multidisciplinary nature of respiratory care, fostering international collaboration is crucial. Research efforts should prioritize cross-border partnerships to ensure that findings are generalizable across different health care settings and patient populations.

Limitations

While this bibliometric analysis provides a comprehensive overview of research trends, it is not without limitations. First, the analysis was limited to publications indexed in the WoSCC, which may have excluded relevant studies from other databases such as Scopus, PubMed, and Embase. This could potentially lead to an underrepresentation of research from regions or countries where publication in these databases are more prevalent.

Second, the exclusion of non-English publications may have limited insights into research conducted in non-English-speaking regions, particularly in countries like China, Japan, and India, which are emerging contributors to prone positioning research. The reliance on English-language publications could introduce a bias, potentially underestimating the global scope of research in this field. Third, bibliometric analyses are inherently quantitative and do not assess the quality or clinical impact of the studies themselves. While citation metrics provide a proxy for influence, they do not always reflect the practical applicability of the research findings in clinical settings. High citation counts may sometimes be driven by factors unrelated to the study's scientific rigor, such as controversies or methodological critiques.

Conclusions

This bibliometric analysis highlights the expanding body of research on prone positioning in respiratory failure, with significant contributions from leading countries, institutions, and authors. The findings underscore the critical role of prone positioning in managing respiratory failure and improving patient outcomes, particularly in the context of the COVID-19 pandemic. Although significant progress has been made in this field, key gaps remain, particularly in understanding the application of prone positioning in nonintubated patients and the long-term outcomes. Future research should focus on addressing these gaps by exploring personalized treatment approaches, fostering international collaboration, and further investigating the physiological mechanisms underlying prone positioning. By building upon the existing evidence base, the medical community can ensure that prone positioning continues to evolve as a more effective and widely adopted intervention in the management of respiratory failure.

XSL•FO
RenderX

Acknowledgments

This work was supported by Yunnan Province Longyun Expert Workstation (202305AF150090). During the preparation of this work, the authors used DeepSeek to polish the language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data Availability

The dataset supporting the conclusions of this article is available in the Web of Science Core Collection (WoSCC) [74]. The analysis code is available on GitHub [75]. The data are available upon request by contacting the corresponding author.

Authors' Contributions

RL, FY, CFY, and MY designed the study. ZHZ and XH conducted the database search using the search strategy and identified the appropriate records for bibliometric analysis. QLL and ZGY performed the Biblioshiny analysis, while RL and KYZ performed the CiteSpace analysis. All authors participated in interpreting the results. RL, FY, CFY, and MY drafted the manuscript with input from all authors. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sheets.
[XLSX File, 68 KB - i-jmr_v14i1e67276_app1.xlsx]

References

1. Fan E, Del Sorbo L, Goligher EC, et al. An Official American Thoracic Society/European Society of Intensive Care Medicine/Society of Critical Care Medicine Clinical Practice Guideline: Mechanical Ventilation in Adult Patients with Acute Respiratory Distress Syndrome. Am J Respir Crit Care Med 2017 May 1;195(9):1253-1263. [doi: 10.1164/rccm.201703-0548ST] [Medline: 28459336]
2. Douglas WW, Rehder K, Beynen FM, Sessler AD, Marsh HM. Improved oxygenation in patients with acute respiratory failure: the prone position. Am Rev Respir Dis 1977 Apr;115(4):559-566. [doi: 10.1164/arrd.1977.115.4.559] [Medline: 322557]
3. Mellins RB. Pulmonary physiotherapy in the pediatric age group. Am Rev Respir Dis 1974 Dec;110(6 Pt 2):137-142. [doi: 10.1164/arrd.1974.110.6P2.137] [Medline: 4613219]
4. Bryan AC. Conference on the scientific basis of respiratory therapy. Pulmonary physiotherapy in the pediatric age group. Comments of a devil's advocate. Am Rev Respir Dis 1974 Dec;110(6 Pt 2):143-144. [doi: 10.1164/arrd.1974.110.6P2.143] [Medline: 4440945]
5. Gattinoni L, Tognoni G, Pesenti A, et al. Effect of prone positioning on the survival of patients with acute respiratory failure. N Engl J Med 2001 Aug 23;345(8):568-573. [doi: 10.1056/NEJMoa010043] [Medline: 11529210]
6. Guérin C, Reignier J, Richard JC, et al. Prone positioning in severe acute respiratory distress syndrome. N Engl J Med 2013 Jun 6;368(23):2159-2168. [doi: 10.1056/NEJMoa1214103] [Medline: 23688302]
7. Li J, Luo J, Pavlov I, et al. Awake prone positioning for non-intubated patients with COVID-19-related acute hypoxaemic respiratory failure: a systematic review and meta-analysis. Lancet Respir Med 2022 Jun;10(6):573-583. [doi: 10.1016/S2213-2600(22)00043-1] [Medline: 35305308]
8. Fossali T, Pavlovsky B, Ottolina D, et al. Effects of Prone Position on Lung Recruitment and Ventilation-Perfusion Matching in Patients With COVID-19 Acute Respiratory Distress Syndrome: A Combined CT Scan/Electrical Impedance Tomography Study. Crit Care Med 2022 May 1;50(5):723-732. [doi: 10.1097/CCM.0000000000005450] [Medline: 35200194]
9. Guérin C, Albert RK, Beitler J, et al. Prone position in ARDS patients: why, when, how and for whom. Intensive Care Med 2020 Dec;46(12):2385-2396. [doi: 10.1007/s00134-020-06306-w]
10. Ahmad P, Slots J. A bibliometric analysis of periodontology. Periodontol 2000 2021 Feb;85(1):237-240. [doi: 10.1111/prd.12376]
11. Zhiguo F, Ji W, Shenyuan C, et al. A swift expanding trend of extracellular vesicles in spinal cord injury research: A bibliometric analysis. J Nanobiotechnology 2023 Aug 23;21(1):289. [doi: 10.1186/s12951-023-02051-6] [Medline: 37612689]
12. Zhang L, Mei S, Zhu B, Zhao Z. Trends in research on acute lung injury/acute respiratory distress syndrome associated with viral pneumonia from 1992 to 2022: a 31-year bibliometric analysis. Front Med 2023;10:37359015. [doi: 10.3389/fmed.2023.1158519]
13. Lu Y, Li W, Qi S, Cheng K, Wu H. Mapping knowledge structure and emerging trends of extracorporeal membrane oxygenation for acute respiratory distress syndrome: a bibliometric and visualized study. Front Med (Lausanne) 2024;11:1365864. [doi: 10.3389/fmed.2024.1365864] [Medline: 39086955]

XSL•FO
RenderX

14. Aria M, Cuccurullo C. bibliometrix : An R-tool for comprehensive science mapping analysis. J Informetr 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]
15. Ping Q, He J, Chen C. How many ways to use CiteSpace? A study of user interactive events over 14 months. Asso Info Sci Tech 2017 May;68(5):1234-1256. [doi: 10.1002/asi.23770]
16. Akwu NA, Lekhooa M, Deqiang D, Aremu AO. Antidepressant effects of coumarins and their derivatives: A critical analysis of research advances. Eur J Pharmacol 2023 Oct 5;956:175958. [doi: 10.1016/j.ejphar.2023.175958] [Medline: 37543158]
17. Thompson DF, Callen EC, Nahata MC. New indices in scholarship assessment. Am J Pharm Educ 2009 Oct 1;73(6):111. [doi: 10.5688/aj7306111] [Medline: 19885080]
18. Li X, Wei W, Wang Y, Wang Q, Liu Z. Global Trend in the Research and Development of Acupuncture Treatment on Parkinson's Disease From 2000 to 2021: A Bibliometric Analysis. Front Neurol 2022;13:906317. [doi: 10.3389/fneur.2022.906317] [Medline: 35873762]
19. Ou P, Wen R, Shi L, Wang J, Liu C. Artificial intelligence empowering rare diseases: a bibliometric perspective over the last two decades. Orphanet J Rare Dis 2024 Sep 13;19(1):345. [doi: 10.1186/s13023-024-03352-1] [Medline: 39272071]
20. Lamm WJ, Graham MM, Albert RK. Mechanism by which the prone position improves oxygenation in acute lung injury. Am J Respir Crit Care Med 1994 Jul;150(1):184-193. [doi: 10.1164/ajrccm.150.1.8025748] [Medline: 8025748]
21. Albert RK. Prone ventilation. Clin Chest Med 2000 Sep;21(3):511-517. [doi: 10.1016/s0272-5231(05)70162-3] [Medline: 11019723]
22. Albert RK, Hubmayr RD. The prone position eliminates compression of the lungs by the heart. Am J Respir Crit Care Med 2000 May;161(5):1660-1665. [doi: 10.1164/ajrccm.161.5.9901037] [Medline: 10806172]
23. Michelet P, Roch A, Gainnier M, Sainty JM, Auffray JP, Papazian L. Influence of support on intra-abdominal pressure, hepatic kinetics of indocyanine green and extravascular lung water during prone positioning in patients with ARDS: a randomized crossover study. Crit Care 2005 Jun;9(3):R251-R257. [doi: 10.1186/cc3513] [Medline: 15987398]
24. Papazian L, Bregeon F, Gaillat F, et al. Respective and combined effects of prone position and inhaled nitric oxide in patients with acute respiratory distress syndrome. Am J Respir Crit Care Med 1998 Feb;157(2):580-585. [doi: 10.1164/ajrccm.157.2.9705046] [Medline: 9476876]
25. Ehrmann S, Li J, Ibarra-Estrada M, et al. Awake prone positioning for COVID-19 acute hypoxaemic respiratory failure: a randomised, controlled, multinational, open-label meta-trial. Lancet Respir Med 2021 Dec;9(12):1387-1395. [doi: 10.1016/S2213-2600(21)00356-8] [Medline: 34425070]
26. Cammarota G, Rossi E, Vitali L, et al. Effect of awake prone position on diaphragmatic thickening fraction in patients assisted by noninvasive ventilation for hypoxemic acute respiratory failure related to novel coronavirus disease. Crit Care 2021 Aug 24;25(1):305. [doi: 10.1186/s13054-021-03735-x] [Medline: 34429131]
27. Gattinoni L, Pesenti A, Carlesso E. Body position changes redistribute lung computed-tomographic density in patients with acute respiratory failure: impact and clinical fallout through the following 20 years. Intensive Care Med 2013 Nov;39(11):1909-1915. [doi: 10.1007/s00134-013-3066-x] [Medline: 24026295]
28. Bellani G, Laffey JG, Pham T, et al. Epidemiology, Patterns of Care, and Mortality for Patients With Acute Respiratory Distress Syndrome in Intensive Care Units in 50 Countries. JAMA 2016 Feb 23;315(8):788-800. [doi: 10.1001/jama.2016.0291] [Medline: 26903337]
29. Rampon GL, Simpson SQ, Agrawal R. Prone Positioning for Acute Hypoxemic Respiratory Failure and ARDS: A Review. Chest 2023 Feb;163(2):332-340. [doi: 10.1016/j.chest.2022.09.020] [Medline: 36162482]
30. McNicholas BA, Ibarra-Estrada M, Perez Y, et al. Awake prone positioning in acute hypoxaemic respiratory failure. Eur Respir Rev 2023 Jun 30;32(168):220245. [doi: 10.1183/16000617.0245-2022] [Medline: 37137508]
31. Voggenreiter G, Neudeck F, Aufmkolk M, et al. Intermittent prone positioning in the treatment of severe and moderate posttraumatic lung injury. Crit Care Med 1999 Nov;27(11):2375-2382. [doi: 10.1097/00003246-199911000-00009] [Medline: 10579251]
32. Stocker R, Neff T, Stein S, Ecknauer E, Trentz O, Russi E. Prone postioning and low-volume pressure-limited ventilation improve survival in patients with severe ARDS. Chest 1997 Apr;111(4):1008-1017. [doi: 10.1378/chest.111.4.1008] [Medline: 9106582]
33. Kornecki A, Frndova H, Coates AL, Shemie SD. 4A randomized trial of prolonged prone positioning in children with acute respiratory failure. Chest 2001 Jan;119(1):211-218. [doi: 10.1378/chest.119.1.211] [Medline: 11157606]
34. Gouna G, Rakza T, Kuissi E, Pennaforte T, Mur S, Storme L. Positioning effects on lung function and breathing pattern in premature newborns. J Pediatr 2013 Jun;162(6):1133-1137. [doi: 10.1016/j.jpeds.2012.11.036] [Medline: 23312684]
35. Prud'homme E, Trigui Y, Elharrar X, et al. Effect of Prone Positioning on the Respiratory Support of Nonintubated Patients With COVID-19 and Acute Hypoxemic Respiratory Failure: A Retrospective Matching Cohort Study. Chest 2021 Jul;160(1):85-88. [doi: 10.1016/j.chest.2021.01.048] [Medline: 33516704]
36. Halpin DMG, Criner GJ, Papi A, et al. Global Initiative for the Diagnosis, Management, and Prevention of Chronic Obstructive Lung Disease. The 2020 GOLD Science Committee Report on COVID-19 and Chronic Obstructive Pulmonary Disease. Am J Respir Crit Care Med 2021 Jan 1;203(1):24-36. [doi: 10.1164/rccm.202009-3533SO] [Medline: 33146552]
37. Guerin C, Gaillard S, Lemasson S, et al. Effects of systematic prone positioning in hypoxemic acute respiratory failure: a randomized controlled trial. JAMA 2004 Nov 17;292(19):2379-2387. [doi: 10.1001/jama.292.19.2379] [Medline: 15547166]

XSL•FO
RenderX

38. Liu L, Sun Q, Zhao H, et al. Prolonged vs shorter awake prone positioning for COVID-19 patients with acute respiratory failure: a multicenter, randomised controlled trial. Intensive Care Med 2024 Aug;50(8):1298-1309. [doi: 10.1007/s00134-024-07545-x] [Medline: 39088076]
39. Albert RK, Leasa D, Sanderson M, Robertson HT, Hlastala MP. The prone position improves arterial oxygenation and reduces shunt in oleic-acid-induced acute lung injury. Am Rev Respir Dis 1987 Mar;135(3):628-633. [doi: 10.1164/arrd.1987.135.3.628] [Medline: 3030168]
40. Papazian L, Roch A, Bregeon F, et al. Inhaled nitric oxide and vasoconstrictors in acute respiratory distress syndrome. Am J Respir Crit Care Med 1999 Aug;160(2):473-479. [doi: 10.1164/ajrccm.160.2.9809110] [Medline: 10430716]
41. Papazian L, Paladini MH, Bregeon F, et al. Is a short trial of prone positioning sufficient to predict the improvement in oxygenation in patients with acute respiratory distress syndrome? Intensive Care Med 2001 Jun;27(6):1044-1049. [doi: 10.1007/s001340000799] [Medline: 11497137]
42. Papazian L, Paladini MH, Bregeon F, et al. Can the tomographic aspect characteristics of patients presenting with acute respiratory distress syndrome predict improvement in oxygenation-related response to the prone position? Anesthesiology 2002 Sep;97(3):599-607. [doi: 10.1097/00000542-200209000-00013] [Medline: 12218526]
43. Langer M, Mascheroni D, Marcolin R, Gattinoni L. The prone position in ARDS patients. A clinical study. Chest 1988 Jul;94(1):103-107. [doi: 10.1378/chest.94.1.103] [Medline: 3383620]
44. Gattinoni L, Pesenti A. The concept of “baby lung”. Intensive Care Med 2005 Jun;31(6):776-784. [doi: 10.1007/s00134-005-2627-z] [Medline: 15812622]
45. Gattinoni L, Vagginelli F, Carlesso E, et al. Decrease in PaCO2 with prone position is predictive of improved outcome in acute respiratory distress syndrome. Crit Care Med 2003 Dec;31(12):2727-2733. [doi: 10.1097/01.CCM.0000098032.34052.F9] [Medline: 14668608]
46. Gattinoni L, Carlesso E, Cadringher P, Valenza F, Vagginelli F, Chiumello D. Physical and biological triggers of ventilator-induced lung injury and its prevention. Eur Respir J Suppl 2003 Nov;47:15s-25s. [doi: 10.1183/09031936.03.00021303] [Medline: 14621113]
47. Pelosi P, D’Onofrio D, Chiumello D, et al. Pulmonary and extrapulmonary acute respiratory distress syndrome are different. Eur Respir J Suppl 2003 Aug;42:48s-56s. [doi: 10.1183/09031936.03.00420803] [Medline: 12946001]
48. Sud S, Friedrich JO, Adhikari NKJ, et al. Effect of prone positioning during mechanical ventilation on mortality among patients with acute respiratory distress syndrome: a systematic review and meta-analysis. CMAJ 2014 Jul 8;186(10):E381-E390. [doi: 10.1503/cmaj.140081] [Medline: 24863923]
49. Grasselli G, Calfee CS, Camporota L, et al. ESICM guidelines on acute respiratory distress syndrome: definition, phenotyping and respiratory support strategies. Intensive Care Med 2023 Jul;49(7):727-759. [doi: 10.1007/s00134-023-07050-7] [Medline: 37326646]
50. Vianello A, Turrin M, Guarnieri G, et al. Prone Positioning Is Safe and May Reduce the Rate of Intubation in Selected COVID-19 Patients Receiving High-Flow Nasal Oxygen Therapy. J Clin Med 2021 Jul 30;10(15):3404. [doi: 10.3390/jcm10153404] [Medline: 34362185]
51. Girard R, Baboi L, Ayzac L, Richard JC, Guérin C, Proseva trial group. The impact of patient positioning on pressure ulcers in patients with severe ARDS: results from a multicentre randomised controlled trial on prone positioning. Intensive Care Med 2014 Mar;40(3):397-403. [doi: 10.1007/s00134-013-3188-1] [Medline: 24352484]
52. Scaramuzzo G, Gamberini L, Tonetti T, et al. Sustained oxygenation improvement after first prone positioning is associated with liberation from mechanical ventilation and mortality in critically ill COVID-19 patients: a cohort study. Ann Intensive Care 2021 Apr 26;11(1):63. [doi: 10.1186/s13613-021-00853-1] [Medline: 33900484]
53. Schmidt M, Zogheib E, Rozé H, et al. The PRESERVE mortality risk score and analysis of long-term outcomes after extracorporeal membrane oxygenation for severe acute respiratory distress syndrome. Intensive Care Med 2013 Oct;39(10):1704-1713. [doi: 10.1007/s00134-013-3037-2] [Medline: 23907497]
54. COVID-ICU Group on behalf of the REVA Network and the COVID-ICU Investigators. Clinical characteristics and day-90 outcomes of 4244 critically ill adults with COVID-19: a prospective cohort study. Intensive Care Med 2021 Jan;47(1):60-73. [doi: 10.1007/s00134-020-06294-x]
55. Vieillard-Baron A, Charron C, Caille V, Belliard G, Page B, Jardin F. Prone positioning unloads the right ventricle in severe ARDS. Chest 2007 Nov;132(5):1440-1446. [doi: 10.1378/chest.07-1013] [Medline: 17925425]
56. Bein T, Bischoff M, Brückner U, et al. Short version S2e guidelines: “Positioning therapy and early mobilization for prophylaxis or therapy of pulmonary function disorders”. Anaesthesist 2015 Aug;64(8):596-611. [doi: 10.1007/s00101-015-0060-4] [Medline: 26260196]
57. Hu SL, He HL, Pan C, et al. The effect of prone positioning on mortality in patients with acute respiratory distress syndrome: a meta-analysis of randomized controlled trials. Crit Care 2014 May 28;18(3):R109. [doi: 10.1186/cc13896] [Medline: 24887034]
58. Tang R, Huang Y, Chen Q, et al. Relationship between regional lung compliance and ventilation homogeneity in the supine and prone position. Acta Anaesthesiol Scand 2012 Oct;56(9):1191-1199. [doi: 10.1111/j.1399-6576.2012.02741.x] [Medline: 22834873]

XSL•FO
RenderX

59. Mao Y, Zhao K, Chen N, et al. A 2-decade bibliometric analysis of epigenetics of cardiovascular disease: from past to present. Clin Epigenetics 2023 Nov 25;15(1):184. [doi: 10.1186/s13148-023-01603-9] [Medline: 38007493]
60. Walther SM, Domino KB, Glenny RW, Hlastala MP. Positive end-expiratory pressure redistributes perfusion to dependent lung regions in supine but not in prone lambs. Crit Care Med 1999 Jan;27(1):37-45. [doi: 10.1097/00003246-199901000-00024] [Medline: 9934891]
61. Broccard A, Shapiro RS, Schmitz LL, Adams AB, Nahum A, Marini JJ. Prone positioning attenuates and redistributes ventilator-induced lung injury in dogs. Crit Care Med 2000 Feb;28(2):295-303. [doi: 10.1097/00003246-200002000-00001] [Medline: 10708156]
62. Mancebo J, Fernández R, Blanch L, et al. A multicenter trial of prolonged prone ventilation in severe acute respiratory distress syndrome. Am J Respir Crit Care Med 2006 Jun 1;173(11):1233-1239. [doi: 10.1164/rccm.200503-353OC] [Medline: 16556697]
63. Mekontso Dessap A, Boissier F, Charron C, et al. Acute cor pulmonale during protective ventilation for acute respiratory distress syndrome: prevalence, predictors, and clinical impact. Intensive Care Med 2016 May;42(5):862-870. [doi: 10.1007/s00134-015-4141-2] [Medline: 26650055]
64. Shelhamer MC, Wesson PD, Solari IL, et al. Prone Positioning in Moderate to Severe Acute Respiratory Distress Syndrome Due to COVID-19: A Cohort Study and Analysis of Physiology. J Intensive Care Med 2021 Feb;36(2):241-252. [doi: 10.1177/0885066620980399] [Medline: 33380236]
65. Douglas IS, Rosenthal CA, Swanson DD, et al. Safety and Outcomes of Prolonged Usual Care Prone Position Mechanical Ventilation to Treat Acute Coronavirus Disease 2019 Hypoxemic Respiratory Failure. Crit Care Med 2021 Mar 1;49(3):490-502. [doi: 10.1097/CCM.0000000000004818] [Medline: 33405409]
66. Schmidt M, Hajage D, Lebreton G, et al. Prone Positioning During Extracorporeal Membrane Oxygenation in Patients With Severe ARDS: The PRONECMO Randomized Clinical Trial. JAMA 2023 Dec 26;330(24):2343-2353. [doi: 10.1001/jama.2023.24491] [Medline: 38038395]
67. Rosén J, von Oelreich E, Fors D, et al. Awake prone positioning in patients with hypoxemic respiratory failure due to COVID-19: the PROFLO multicenter randomized clinical trial. Crit Care 2021 Jun 14;25(1):209. [doi: 10.1186/s13054-021-03602-9] [Medline: 34127046]
68. Bhandari AP, Nnate DA, Vasanthan L, Konstantinidis M, Thompson J. Positioning for acute respiratory distress in hospitalised infants and children. Cochrane Database Syst Rev 2022 Jun 6;6(6):CD003645. [doi: 10.1002/14651858.CD003645.pub4] [Medline: 35661343]
69. Xin Y, Cereda M, Hamedani H, et al. Positional Therapy and Regional Pulmonary Ventilation. Anesthesiology 2020 Nov 1;133(5):1093-1105. [doi: 10.1097/ALN.0000000000003509] [Medline: 32773690]
70. Walter T, Ricard JD. Extended prone positioning for intubated ARDS: a review. Crit Care 2023 Jul 5;27(1):264. [doi: 10.1186/s13054-023-04526-2] [Medline: 37408074]
71. De Jong A, Wrigge H, Hedenstierna G, et al. How to ventilate obese patients in the ICU. Intensive Care Med 2020 Dec;46(12):2423-2435. [doi: 10.1007/s00134-020-06286-x] [Medline: 33095284]
72. Roldán R, Rodriguez S, Barriga F, et al. Sequential lateral positioning as a new lung recruitment maneuver: an exploratory study in early mechanically ventilated Covid-19 ARDS patients. Ann Intensive Care 2022 Feb 12;12(1):13. [doi: 10.1186/s13613-022-00988-9] [Medline: 35150355]
73. Bouchant L, Godet T, Arpajou G, et al. Physiological effects and safety of bed verticalization in patients with acute respiratory distress syndrome. Crit Care 2024 Aug 5;28(1):262. [doi: 10.1186/s13054-024-05013-y] [Medline: 39103928]
74. Clarivate. Web of Science platform. URL: https://clarivate.com/academia-government/scientific-and-academic-research/research-discovery-and-referencing/web-of-science/ [accessed 2025-06-20]
75. GitHub. leoqj/biblioshiny. URL: https://github.com/leoqj/biblioshiny [accessed 2025-06-20]

Abbreviations

ARDS: Acute Respiratory Distress Syndrome
ARF: Acute Respiratory Failure
COVID-19: coronavirus disease 2019
MCP: Multiple-Country Publications
WoSCC: Web of Science Core Collection

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 07.10.24; peer-reviewed by X Meng, Y Zeng; revised version received 31.03.25; accepted 14.04.25; published 20.06.25.

Please cite as:
Lei R, Yue F, Yue C, Zhang Z, Huang X, Li Q, Yang Z, Li R, Zhao K, Yang M
Characterization of Global Research Trends and Prospects on Prone Positioning in Respiratory Failure: Bibliometric Analysis
Interact J Med Res 2025;14:e67276
URL: https://www.i-jmr.org/2025/1/e67276
doi:10.2196/67276

© Rong Lei, Feng Yue, Chaofu Yue, Zihan Zhang, Xian Huang, Qiaolin Li, Zhigang Yang, Rong Li, Keyi Zhao, Mei Yang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 20.6.2025. 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

Lactation Consultant Access and Breastfeeding Outcomes in the United States: Cross-Sectional Analysis

James F Dockins[1*], EdD; Heather D Pahl[2*], MHCA; David J Lingerfelt[1*], EdD

[1]Rockhurst University, 1100 Rockhurst Road, 300 Conway Hall, Kansas City, MO, United States

[2]University of Arkansas – Fort Smith, Fort Smith, AR, United States

[*]all authors contributed equally

Corresponding Author:
James F Dockins, EdD
Rockhurst University, 1100 Rockhurst Road, 300 Conway Hall, Kansas City, MO, United States

Abstract

Background: Breastfeeding provides unmatched health, developmental, and economic benefits to both infants and mothers, yet breastfeeding continuation rates remain suboptimal in the United States, especially beyond the early postpartum period. Despite well-documented advantages, many mothers face challenges that lead to early cessation, including lack of access to skilled lactation support. International Board Certified Lactation Consultants (IBCLCs) are considered the clinical gold standard in lactation care, but their availability varies widely across states. Understanding how IBCLC access relates to breastfeeding outcomes at the population level is critical to informing equitable public health interventions.

Objective: The aim of this study is to determine whether state-level IBCLC density is associated with breastfeeding initiation and exclusive breastfeeding at 3 and 6 months.

Methods: This cross-sectional analysis used publicly available 2022 data from the Centers for Disease Control and Prevention, US Census Bureau, and the International Board of Lactation Consultant Examiners. IBCLC density per 100,000 women of childbearing age (15 - 49 years) was calculated for each of the 50 US states. Breastfeeding outcome data included initiation, exclusive breastfeeding at 3 months, and exclusive breastfeeding at 6 months. Simple and multiple linear regressions were conducted to evaluate the association between IBCLC density and breastfeeding outcomes, adjusting for income, education, and insurance coverage.

Results: IBCLC density ranged from 14.4 to 60.7 per 100,000 women of childbearing age across US states, with a national average of 25.5. Pearson correlation analysis indicated significant positive associations between IBCLC density and breastfeeding outcomes, including initiation (r=0.38; P<.001), exclusive breastfeeding at 3 months (r=0.52; P<.001), and exclusive breastfeeding at 6 months (r=0.32; P<.001). In multiple linear regression models adjusting for income, education, and insurance status, IBCLC density remained significantly associated with all 3 outcomes. For breastfeeding initiation, the adjusted β was 0.26 (95% CI 0.08-0.44; P=.005); for exclusive breastfeeding at 3 months, β was 0.43 (95% CI 0.23-0.63; P<.001); and for exclusive breastfeeding at 6 months, β was 0.25 (95% CI 0.12-0.39; P<.001). Adjusted R^2 values for the models ranged from 0.42 to 0.44, indicating moderate explanatory power.

Conclusions: Higher IBCLC density is significantly associated with improved breastfeeding outcomes at the state level, particularly exclusive breastfeeding at 3 months. These findings support initiatives to expand access to professional lactation support as part of public health strategies to improve breastfeeding rates.

(Interact J Med Res 2025;14:e70098) doi:10.2196/70098

KEYWORDS

breastfeeding; lactation consultant; exclusive breastfeeding; artificial baby milk

Introduction

Background

Breastfeeding is a complex, symbiotic process between a mother and child and breast milk is the one source of infant nutrition that is perfectly, biologically suited for an infant in its unadulterated form. The value of breastfeeding goes well beyond meeting nutritional needs and it is a mother's way to continue her uniquely profound role of nourishing and nurturing after childbirth. Although breastfeeding is the biological norm for humans, it is no longer the social norm. With the early 20th century commercialization of infant formula, which produces over US $55 billion in sales annually, mothers have been dissuaded from breastfeeding [1]. Infant formula marketing is partially to blame for the decrease in breastfeeding on a global scale—a drop from 90% in the 20th century to 42% in the 21st century [2]. Infant formula usage increasingly became the social

XSL•FO
RenderX

norm while mothers increasingly lost their innate ability to breastfeed. Mothers have fewer supporters of breastfeeding as fewer people breastfeed, including their role models and peers. Although breastfeeding is innate, it can be incredibly challenging, especially considering most mothers today do not grow up seeing others breastfeed regularly. The potential problems that mothers can encounter are numerous and often lead to a mother discontinuing breastfeeding prematurely [3].

In the United States, approximately 83% of infants have ever breastfed [4]. This figure even includes babies who have latched just once in their life after birth, but were never breastfed again. By 3 months, the proportion of infants exclusively breastfed drops to approximately 45%, and then to 25% at 6 months. Nearly one-fifth of breastfed infants have already been given infant formula (artificial baby milk) before they are 2 days old. In a study by Perrine et al [5], approximately 68% of the participants did not meet their intended breastfeeding goal. The overarching message is that most women are breastfeeding for shorter durations than they intended.

Lactation consultants are clinical health care professionals that make it their mission to help mothers meet their breastfeeding goals and make breastfeeding the social norm again.

Lactation consultants are extensively educated and trained to understand the biological, anatomical, physiological, psychological, and social mechanisms of breastfeeding—on the part of both the mother and the infant. They are typically employed by hospitals or pediatric clinics or work in private practice. Lactation consultants are effective at improving breastfeeding outcomes, including exclusivity, duration, self-efficacy, and maternal mental health [6,7]. Although lactation consultants do indeed solve many breastfeeding problems, the problem of access to care remains. The lactation consultant profession is fairly small (approximately 19,000 in the United States and 35,000 worldwide) but growing, and it can be difficult for a mother to have geographical access to a lactation consultant [8].

Research Questions and Hypotheses

This study aims to evaluate the relationship between the population density of lactation consultants and breastfeeding outcomes across US states.

Research question: Is there a correlation between the population density of lactation consultants and breastfeeding outcomes?

The H (null hypothesis) is that there is no correlation between the population density of lactation consultants and breastfeeding outcomes. The H_1 (alternative hypothesis) is that there is a correlation between the population density of lactation consultants and breastfeeding outcomes.

Study Rationale

Given that breastfeeding rates nationally and globally are low, especially for continuation of breastfeeding, the focus of this article examines access to lactation consultants, who are designated clinical breastfeeding professionals. Access to lactation consultants is a topic that has not been well researched. In the United States, postnatal support varies significantly depending on geographic location, health care provider networks, and insurance status. Unlike some countries with nationalized health care systems, the United States lacks a standardized postnatal care pathway, and support for breastfeeding often depends on fragmented services offered through hospitals, outpatient clinics, and community health programs. However, there is research indicating that the use of lactation consultants does contribute to better breastfeeding outcomes, including longer duration and fewer problems. The inclusion of key sociodemographic variables in this study—household income, education level, and health insurance coverage—reflects their importance as social determinants of health. These factors can significantly influence a mother's ability to access lactation support services. For example, lower-income mothers or those without insurance may face barriers such as out-of-pocket costs or limited availability of lactation consultants in their communities. Similarly, differences in educational attainment may affect awareness of available resources and breastfeeding best practices. One gap in the current literature is an examination of the relationship between the access to lactation consultants and breastfeeding outcomes. Access to lactation consultants is a different problem entirely than the effectiveness of lactation consultants.

The objective of this study is to examine access to lactation consultants in the United States by measuring their population density at the state level and analyzing its association with key breastfeeding outcomes.

Subsequently, if the results indicate a higher population density of lactation consultants (International Board Certified Lactation Consultants [IBCLCs]) in an area results in better breastfeeding rates, a case can be made for public policy and funding to support improved access to lactation consultants.

Methods

Study Design and Setting

This study used a cross-sectional ecological design to investigate associations between state-level availability of lactation consultants and breastfeeding outcomes across the 50 US states. Data were compiled from publicly available national sources representing the year 2022. No human participants were recruited, as the analysis was based on aggregated secondary data. The setting spanned all 50 US states, and data collection periods correspond to annual releases from the US Census Bureau, the Centers for Disease Control and Prevention (CDC), and the International Board of Lactation Consultant Examiners (IBLCE).

Participants and Eligibility

The unit of analysis was each individual US state. All 50 states were eligible and included in the study based on the availability of complete data across key variables. The District of Columbia and US territories were excluded to ensure consistency in health care infrastructure, population metrics, and reporting standards. As an ecological study, eligibility criteria were based on data completeness at the state level rather than individual characteristics.

XSL•FO
RenderX

Data Sources and Assessments

Workforce statistics for IBCLCs were obtained from IBLCE's published reports, which detail the number of actively certified professionals per state. IBCLC density, the primary exposure variable, was calculated as the number of IBCLCs per 100,000 women aged 15-49 years using 2022 US Census Bureau data [9,10]. Breastfeeding outcomes were obtained from the CDC's 2022 Breastfeeding Report Card, which draws on data from the National Immunization Survey. The 3 key outcomes assessed were breastfeeding initiation (defined as ever having breastfed), exclusive breastfeeding at 3 months, and exclusive breastfeeding at 6 months. These outcomes represent population-level estimates based on parent-reported data and are considered valid for surveillance purposes.

The Breastfeeding Report Card data are derived from the National Immunization Survey, a telephone-based survey of parents with children aged 19-35 months. The CDC uses standardized weighting methods to ensure national representativeness. Although this is a well-established source of breastfeeding surveillance data, it is important to acknowledge that self-reported data can be subject to recall and social desirability bias.

Additionally, the study used the World Health Organization's definition of women of reproductive age—15 to 49 years old—to calculate IBCLC density [11]. This aligns with global demographic standards and ensures comparability across public health data sources.

The study also included 3 covariates identified as potential confounders based on existing literature about social determinants of health: average household income, the percentage of the population without health insurance, and the percentage of women aged 25 years and older with a bachelor's degree or higher. These variables were extracted from the US Census Bureau's 2022 state-level data. Together, these measures were used to assess sociodemographic context and its potential influence on breastfeeding behaviors.

Study Size

The final study population included all 50 US states. This sample size was determined by the availability of complete and consistent state-level data for all variables of interest. As a fixed and exhaustive sample of all states, a formal power analysis or sample size calculation was not necessary or applicable to this ecological analysis.

Data Analysis

Descriptive statistics were used to summarize the distribution of IBCLC density and breastfeeding outcomes across states. Simple linear regression models were conducted to assess the unadjusted associations between IBCLC density and each of the 3 breastfeeding outcomes. Subsequently, multiple linear regression models were developed to control for the influence of income, insurance coverage, and educational attainment. The primary independent variable in all models was IBCLC density, while the dependent variables were breastfeeding initiation, exclusive breastfeeding at 3 months, and exclusive breastfeeding at 6 months.

Pearson correlation analyses were conducted to examine the strength and direction of linear relationships between IBCLC density and each of the 3 breastfeeding outcomes: initiation, exclusive breastfeeding at 3 months, and exclusive breastfeeding at 6 months. These analyses provided initial insight into bivariate associations and guided the selection of variables for subsequent regression modeling. Correlation coefficients (r) and P values were reported to assess statistical significance.

Regression results were reported as standardized beta coefficients (β) along with 95% CIs, P values, and adjusted R^2 values to assess statistical significance and explanatory power. All analyses were conducted using Python programming tools (version 3.11; Python Software Foundation).

Ethical Considerations

This research involved the secondary analysis of publicly available, deidentified datasets and did not include any interaction with human subjects or use of private, identifiable information. Accordingly, ethics board review was not sought, in accordance with US Department of Health and Human Services regulations (45 CFR 46.102), which state that such studies do not constitute human subjects research. Institutional policy at Rockhurst University confirms that publicly available and anonymized data analyses are exempt from institutional review board review. All data used in this study were obtained from official public sources, including the CDC, US Census Bureau, and IBLCE. The study adhered to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for transparent reporting of observational cross-sectional research.

Statistical Testing

To enable consistent cross-state comparisons, the researchers first calculated the density of IBCLCs per 100,000 females of childbearing age (15 - 49 years) using Microsoft Excel 2021 (Microsoft Corp).

Simple linear regression analyses were conducted in Python to evaluate the unadjusted relationship between IBCLC density and each of the 3 key breastfeeding outcomes:

1. Breastfeeding initiation.
2. Exclusive breastfeeding at 3 months.
3. Exclusive breastfeeding at 6 months.

Multiple linear regression was then used to control for potential confounding variables. The dependent (outcome) variables remained the same: breastfeeding initiation, exclusive breastfeeding at 3 months, and exclusive breastfeeding at 6 months. The independent variables in the adjusted models included:

- IBCLC population density (primary exposure)
- Average household income
- Proportion of the population without health insurance
- Proportion of the population with a bachelor's degree or higher

These covariates were selected based on their theoretical importance as social determinants of health, supported by prior research linking these factors to maternal and infant health

XSL•FO
RenderX

disparities [12]. Variables were added to the regression models simultaneously; no stepwise (forward or backward) selection methods were applied.

For each model, we report regression coefficients, *P* values, and adjusted R^2 values to assess the strength, significance, and explanatory power of the observed relationships.

For reader clarity, we have included definitions of relevant terms in Multimedia Appendix 1.

Results

The relevant demographic makeup for each state is shown in Table 1, along with measures for the United States on average.

The breastfeeding initiation, exclusivity at 3 months, and exclusivity at 6 months rates for each state and the United States averages are shown in Table 2.

XSL•FO
RenderX

Table . Demographics by state.

State	Female childbearing population, n	Income, US $	Bachelor's or higher, %	No health insurance, %
Alabama	1,143,118	59,674	28.8	8.8
Alaska	163,121	88,121	30.6	11
Arizona	1,637,098	74,568	33	10.3
Arkansas	677,271	55,432	25.4	8.4
California	9,124,204	91,551	37	6.5
Colorado	1,380,866	89,302	45.9	7.1
Connecticut	804,136	88,429	41.9	5.2
Delaware	216,513	82,174	36.5	5.6
Florida	4,695,524	69,303	34.3	11.2
Georgia	2,599,616	72,837	34.7	11.7
Hawaii	302,734	92,458	35.4	3.5
Idaho	436,965	72,785	32.3	8.2
Illinois	2,852,991	76,708	37.7	6.6
Indiana	1,530,011	66,785	29.6	7
Iowa	693,875	69,588	32.3	4.5
Kansas	649,589	68,925	35.6	8.6
Kentucky	988,972	59,341	27.9	5.6
Louisiana	1,043,544	55,416	27.1	6.9
Maine	282,427	69,543	36.1	6.6
Maryland	1,395,676	94,991	43.8	6.1
Massachusetts	1,610,759	94,488	46.6	2.4
Michigan	2,176,561	66,986	32.1	4.5
Minnesota	1,255,719	82,338	39.1	4.5
Mississippi	669,121	52,719	24.8	10.8
Missouri	1,368,368	64,811	32.2	8.6
Montana	239,382	67,631	34.6	8.3
Nebraska	435,512	69,597	34.7	6.7
Nevada	719,713	72,333	27	11.1
New Hampshire	290,658	89,992	41.3	4.9
New Jersey	2,039,450	96,346	43.5	6.8
New Mexico	464,654	59,726	30.5	8.2
New York	4,457,006	79,557	40	4.9
North Carolina	2,438,121	67,481	35.9	9.3
North Dakota	172,587	71,970	31.8	6.4
Ohio	2,567,828	65,720	32	5.9
Oklahoma	911,644	59,673	28.5	11.7
Oregon	960,141	75,657	36.3	6
Pennsylvania	2,804,185	71,798	35.1	5.3
Rhode Island	246,284	81,854	39.6	4.2
South Carolina	1,169,623	64,115	32.6	9.1
South Dakota	191,168	69,728	31.6	8.1

XSL•FO
RenderX

State	Female childbearing population, n	Income, US $	Bachelor's or higher, %	No health insurance, %
Tennessee	1,598,739	65,254	31.1	9.3
Texas	7,214,941	72,284	33.9	16.6
Utah	842,091	89,168	37.9	8.1
Vermont	137,539	73,991	44.2	3.9
Virginia	1,978,978	85,873	42.2	6.5
Washington	1,786,627	91,306	39.5	6.1
West Virginia	365,994	54,329	24.8	5.9
Wisconsin	1,269,086	70,996	33.2	5.2
Wyoming	124,653	70,042	29.6	11.5
United States	75,125,383	74,755	35.7	8

XSL•FO
RenderX

Table . Breastfeeding rates by state.

State	Initiation, %	3 months, %	6 months, %
Alabama	71.1	38	21
Alaska	92.9	57.6	30.9
Arizona	85.4	43.2	24
Arkansas	74.9	42	24.4
California	89.9	51.6	27.3
Colorado	94	62.8	32.1
Connecticut	84.2	44.7	26.3
Delaware	83.6	48.3	25
Florida	71	32.4	18.2
Georgia	82.6	39.9	18.7
Hawaii	90.1	50.6	27.7
Idaho	93.5	57.6	30.4
Illinois	84.9	47.8	28.3
Indiana	85.9	46.2	21.5
Iowa	82.4	52.8	27
Kansas	87.1	47	29.2
Kentucky	74.7	35.4	21.2
Louisiana	71.1	38	22.2
Maine	86.6	50.5	28.7
North Dakota	85.7	48.8	27.4
Massachusetts	80	52.8	29.2
Michigan	83.1	42.6	25.1
Minnesota	91.9	57.5	36.5
Mississippi	69.4	31.1	15.6
Missouri	78.3	42.5	24.6
Montana	83.5	50.4	34.3
Nebraska	86.1	49.3	26
Nevada	83.8	42.4	22.3
New Hampshire	82.2	55	31.8
New Jersey	82.5	41.2	23.5
New Mexico	83.4	52.3	29
New York	86.7	42.4	23.4
North Carolina	83.4	47.2	22.1
North Dakota	85.7	48.8	27.4
Ohio	79.5	42.7	23.7
Oklahoma	77.3	43.1	23.2
Oregon	87.2	59.2	34.2
Pennsylvania	74.8	42.4	24.6
Rhode Island	82.4	42.3	22.9
South Carolina	80.6	43.3	19.3
South Dakota	88.9	52.1	29.1
Tennessee	78.8	41.9	24.9

XSL•FO
RenderX

State	Initiation, %	3 months, %	6 months, %
Texas	84.1	42.4	24
Utah	91.4	49.5	27.3
Vermont	91.8	61	36.2
Virginia	83.3	39.6	25.8
Washington	93.7	57	29.5
West Virginia	59.8	28	13.8
Wisconsin	87.5	59.3	31.3
Wyoming	92.4	55.3	27.2
United States	83.2	45.3	24.9

The initiation rates range from 59.8% in West Virginia to 94% in Colorado. Exclusivity rates at 3 months range from 28% in West Virginia to 62.8% in Colorado. Exclusivity rates at 6 months range from 13.8% in West Virginia to 36.5% in Minnesota. The number of IBCLCs and IBCLC density in each state are shown in Table 3.

XSL•FO
RenderX

Table . IBCLC[a] density by state.

State	Total IBCLCs	Female childbearing population	IBCLCs per 100,000
Alabama	211	1,143,118	18.5
Alaska	99	163,121	60.7
Arizona	414	1,637,098	25.3
Arkansas	117	677,271	17.3
California	2640	9,124,204	28.9
Colorado	446	1,380,866	32.3
Connecticut	243	804,136	30.2
Delaware	59	216,513	27.3
Florida	788	4,695,524	16.8
Georgia	515	2,599,616	19.8
Hawaii	101	302,734	33.4
Idaho	135	436,965	30.9
Illinois	624	2,852,991	21.9
Indiana	470	1,530,011	30.7
Iowa	162	693,875	23.3
Kansas	210	649,589	32.3
Kentucky	167	988,972	16.9
Louisiana	195	1,043,544	18.7
Maine	85	282,427	30.1
Maryland	495	1,395,676	35.5
Massachusetts	458	1,610,759	28.4
Michigan	482	2,176,561	22.1
Minnesota	443	1,255,719	35.3
Mississippi	97	669,121	14.5
Missouri	360	1,368,368	26.3
Montana	65	239,382	27.2
Nebraska	144	435,512	33.1
Nevada	108	719,713	15.0
New Hampshire	95	290,658	32.7
New Jersey	514	2,039,450	25.2
New Mexico	135	464,654	29.1
New York	1057	4,457,006	23.7
North Carolina	778	2,438,121	31.9
North Dakota	33	172,587	19.1
Ohio	708	2,567,828	27.6
Oklahoma	210	911,644	23.0
Oregon	510	960,141	53.1
Pennsylvania	726	2,804,185	25.9
Rhode Island	60	246,284	24.4
South Carolina	263	1,169,623	22.5
South Dakota	42	191,168	22.0
Tennessee	312	1,598,739	19.5

XSL•FO
RenderX

State	Total IBCLCs	Female childbearing population	IBCLCs per 100,000
Texas	1401	7,214,941	19.4
Utah	183	842,091	21.7
Vermont	82	137,539	59.6
Virginia	621	1,978,978	31.4
Washington	626	1,786,627	35.0
West Virginia	73	365,994	19.9
Wisconsin	368	1,269,086	29.0
Wyoming	18	124,653	14.4
United States	19,148	75,125,383	25.5

[a]IBCLC: International Board Certified Lactation Consultant.

The average IBCLC density for the entire United States female childbearing age population is 25.5, while the state with the highest density of IBCLCs was Alaska at 60.7 and the lowest density was in Wyoming at 14.4.

The simple linear regression for IBCLC density and breastfeeding initiation shown in Table 4 yielded a correlation coefficient of 0.38 with a *P* value of <.001 and an R^2 of 0.26.

The simple linear regression for IBCLC density and exclusive breastfeeding at 3 months shown in Table 5 yielded a correlation coefficient of 0.52 with a *P* value of <.001 and an R^2 of 0.41.

The simple linear regression for IBCLC density and exclusive breastfeeding at 6 months shown in Table 6 yielded a correlation coefficient of 0.32 with a *P* value of <.001 and an R^2 of 0.41.

The multiple linear regression analysis for breastfeeding initiation shown in Table 7 has an adjusted R^2 of 0.44, and it produced a coefficient of 0.26 for IBCLC density with a *P* value of .005. Income had an extremely low coefficient of 0.0003 with a *P* value of .005. Lack of health insurance coverage yielded a very high coefficient of 47.93, but a *P* value of .14. Having a bachelor's degree or higher yielded a coefficient of −6.05 and *P* value of .82.

Table . Simple linear regression of IBCLC[a] density and breastfeeding initiation.[b]

	Coefficient	SE	*t* test	*P*>|*t*|	95% CI
Constant	72.9925	2.5770	28.3250	<.001	67.8112-78.1739
IBCLCs per 100,000	0.3783	0.0891	4.2468	<.001	0.1992-0.5573

[a]IBCLC: International Board Certified Lactation Consultant.

[b]The parameters of the linear regression are described here. Model: ordinary least squares. Dependent variable: Initiation. Date: October 18, 2023. Number of observations: 50. Degrees of freedom of the model: 1. Residual degrees of freedom: 48. R^2: 0.273. Adjusted R^2: 0.258. Akaike information criterion: 324.9093. Bayesian information criterion: 328.7333. Log-likelihood: −160.45. *F* statistic: 18.04. Probability (*F* statistic): 9.89E-05. Scale: 37.38.

Table . Simple linear regression of IBCLC[a] density and exclusive breastfeeding at 3 months.[b]

	Coefficient	SE	*t* test	*P*>|*t*|	95% CI
Constant	32.8530	2.5610	12.8283	<.001	27.7038-38.0022
IBCLCs per 100,000	0.5189	0.0885	5.8623	<.001	0.3409-0.6969

[a]IBCLC: International Board Certified Lactation Consultant.

[b]The parameters of the linear regression are described here. Model: ordinary least squares. Dependent variable: 3 months. Date: October 18, 2023. Number of observations: 50. Degrees of freedom of the model: 1. Residual degrees of freedom: 48. R^2: 0.417. Adjusted R^2: 0.405. Akaike information criterion: 324.2871. Bayesian information criterion: 328.1112. Log-likelihood: −160.14. *F* statistic: 34.37. Probability (*F* statistic): 4.07E-07. Scale: 36.918.

XSL•FO
RenderX

Table . Simple linear regression of IBCLC[a] density and exclusive breastfeeding at 6 months.[b]

	Coefficient	SE	*t* test	*P*>|*t*|	95% CI
Constant	17.2583	1.5714	10.9830	<.001	14.0989-20.4178
IBCLCs per 100,000	0.3202	0.0543	5.8959	<.001	0.2110-0.4294

[a]IBCLC: International Board Certified Lactation Consultant.

[b]The parameters of the linear regression are described here. Model: ordinary least squares. Dependent variable: 6 months. Date: October 18, 2023. Number of observations: 50. Degrees of freedom of the model: 1. Residual degrees of freedom: 48. R^2: 0.42. Adjusted R^2: 0.408. Akaike information criterion: 275.4432. Bayesian information criterion: 279.2673. Log-likelihood: −135.72. *F* statistic: 34.76. Probability (*F* statistic): 3.61E-07. Scale: 13.899.

Table . Multiple regression for initiation.[a]

	Coefficient	SE	*t* test	*P*>|*t*|	95% CI
Constant	50.1350	6.9303	7.2342	<.001	36.1767 to 64.0934
IBCLCs[b] per 100,000	0.2623	0.0882	2.9742	.005	0.0847 to 0.4399
Income	0.0003	0.0001	2.9213	.005	0.0001 to 0.0006
No health insurance	47.9309	31.9921	1.4982	.14	−16.5045 to 112.3662
Bachelor’s degree and higher	−6.0498	25.9169	−0.2334	.82	−58.2491 to 46.1496

[a]The parameters of the multiple regression are described here. Model: ordinary least squares. Dependent variable: initiation. Date: October 19, 2023. Number of observations: 50. Degrees of freedom of the model: 4. Residual degrees of freedom: 45. R^2: 0.484. Adjusted R^2: 0.438. Akaike information criterion: 313.7569. Bayesian information criterion: 323.317. Log-likelihood: −151.88. *F* statistic: 10.56. Probability (*F* statistic): 4.04E-06. Scale: 28.293.

[b]IBCLC: International Board Certified Lactation Consultant.

The results for exclusive breastfeeding at 3 months shown in Table 8 are as follows: adjusted R^2=0.42, IBCLC density coefficient=0.43 with *P*<.001, income coefficient=0.0002 with *P*=.20, lack of health insurance coefficient=−8.17 with *P*=.82, and bachelor’s degree or higher coefficient=−0.62 with *P*=.98.

Table . Multiple regression for exclusive breastfeeding at 3 months.[a]

	Coefficient	SE	*t* test	*P*>|*t*|	95% CI
Constant	23.8639	7.7936	3.0620	.004	8.1667 to 39.5610
IBCLCs[b] per 100,000	0.4282	0.0992	4.3175	<.001	0.2284 to 0.6279
Income	0.0002	0.0001	1.2987	.20	−0.0001 to 0.0004
No health insurance	−8.1700	35.9774	−0.2271	.82	−80.6322 to 64.2921
Bachelor’s degree or higher	−0.6175	29.1454	−0.0212	.98	−59.3194 to 58.0844

[a]The parameters of the multiple regression are described here. Model: ordinary least squares. Dependent variable: 3 months. Date: October 18, 2023. Number of observations: 50. Degrees of freedom of the model: 4. Residual degrees of freedom: 45. R^2: 0.47. Adjusted R^2: 0.423. Akaike information criterion: 325.4971. Bayesian information criterion: 335.0572. Log-likelihood: −157.75. *F* statistic: 9.996. Probability (*F* statistic): 7.10E-06. Scale: 35.781.

[b]IBCLC: International Board Certified Lactation Consultant.

Exclusive breastfeeding at 6 months results shown in Table 9 are as follows: adjusted R^2=0.44, IBCLC density coefficient=0.25 with *P*<.001, income coefficient=0.0000 with *P*=.74, lack of health insurance coefficient=−15.34 with *P*=.49, and bachelor’s degree or higher coefficient=15.9 with *P*=.37.

XSL•FO
RenderX

Table . Multiple regression for exclusive breastfeeding at 6 months.[a]

	Coefficient	SE	*t* test	*P*>|*t*|	95% CI
Constant	12.7784	4.7405	2.6956	<.001	3.2306 to 22.3263
IBCLCs[b] per 100,000	0.2537	0.0603	4.2054	.001	0.1322 to 0.3752
Income	0.0000	0.0001	0.3343	.74	−0.0001 to 0.0002
No health insurance	−15.3373	21.8834	−0.7009	.49	−59.4127 to 28.7381
Bachelor's degree or higher	15.8958	17.7278	0.8967	.37	−19.8099 to 51.6015

[a]The parameters of the multiple regression are described here. Model: ordinary least squares. Dependent variable: 6 months. Date: October 18, 2023. Number of observations: 50. Degrees of freedom of the model: 4. Residual degrees of freedom: 45. R^2: 0.482. Adjusted R^2: 0.436. Akaike information criterion: 275.7809. Bayesian information criterion: 285.341. Log-likelihood: −132.89. *F* statistic: 9.996. Probability (*F* statistic): 4.40E-06. Scale: 35.7.

[b]IBCLC: International Board Certified Lactation Consultant.

Discussion

Principal Findings

This study found a positive association between the population density of IBCLCs and key breastfeeding outcomes in the United States, including initiation, exclusive breastfeeding at 3 months, and exclusive breastfeeding at 6 months (Multimedia Appendix 2). The strongest relationship was observed for exclusive breastfeeding at 3 months. These findings align with previous research showing that IBCLCs are associated with improved breastfeeding duration and exclusivity. For example, Patel and Patel [7] reported significant increases in breastfeeding rates when lactation support was provided in both hospital and community settings, and our findings reinforce the importance of IBCLC access beyond the point of care delivery. Similarly, Chrzan-Dętkoś et al [6] found that lactation consultations not only improved exclusivity but also maternal mental health, suggesting multifaceted benefits that may partly explain the state-level effects observed in our analysis.

Importantly, our study adds to this body of work by analyzing IBCLC density as a structural factor at the population level. Although Perrine et al [5] showed that hospital practices aligned with the Baby-Friendly Hospital Initiative improve exclusive breastfeeding intentions, our data demonstrate that widespread IBCLC availability across an entire state is similarly associated with improved breastfeeding metrics. This macro-level perspective complements prior individual-level studies and offers a broader lens on access equity.

These findings suggest that increased access to professional lactation support may play an important role in promoting breastfeeding continuation during the critical early months of infancy.

Comparison to Prior Work

The association between IBCLC density and breastfeeding outcomes identified in this study is consistent with literature emphasizing the clinical effectiveness of professional lactation support. Previous studies have documented improvements in breastfeeding duration and maternal confidence following targeted lactation interventions [6,7]. However, these studies were conducted in localized or clinical settings. Our analysis broadens this understanding by showing that even at the population level, the presence of IBCLCs is significantly associated with improved breastfeeding outcomes, indicating that workforce distribution matters as much as clinical quality.

Additionally, Rollins et al [1] emphasized how gaps in health system structures and commercial pressures undermine breastfeeding support globally. Our findings offer a US-specific parallel, suggesting that one way to counteract these systemic barriers is to increase the geographic accessibility of IBCLCs. Unlike initiatives limited to single institutions, the presence of IBCLCs across regions may mitigate inequities in maternal health resources.

Furthermore, while Perrine et al [5] focused on hospital practices, we extend their insights by suggesting that system-level workforce planning (ie, ensuring every region has adequate access to IBCLCs) could serve as a complementary public health strategy. By embedding our results in this broader context, we highlight how both institutional practices and population-based interventions may be needed to achieve national breastfeeding goals.

Strengths and Limitations

This study has several strengths. It draws from comprehensive, publicly available national datasets, applies both simple and multiple linear regression models, and includes key sociodemographic variables as potential confounders. This enhances the reliability of the observed associations and provides a foundation for future public health policy exploration.

However, several limitations must be acknowledged.

First, the study relies on secondary data sources, which limits the inclusion of other relevant variables such as cultural norms, maternity leave policies, or maternal health conditions. Attempts were made to address this through the inclusion of known confounders (income, education, and insurance status), but unmeasured variables likely remain.

Second, the focus on IBCLCs exclusively excludes other types of lactation support providers (eg, Certified Lactation Counselors; Women, Infants, and Children Peer Counselors), which may result in an underestimation of lactation support availability in some regions. This may particularly affect underserved areas where IBCLCs are sparse but alternative supports are present.

XSL•FO
RenderX

Third, the breastfeeding outcome data are based on parent-reported responses collected through telephone surveys, which may be susceptible to recall or social desirability bias. Although the CDC's methods are well-established and designed to minimize such issues, this limitation is inherent in self-reported data.

Fourth, the analysis was conducted at the state level, which may obscure important regional disparities. Breastfeeding behaviors and IBCLC availability likely vary significantly within states, suggesting the need for future studies using data at the county or ZIP code level.

Future Directions

To build on these findings, future research should focus on more granular geographic data, such as IBCLC access by zip code or county, to better identify disparities masked by state-level averages. Linking provider registries with localized breastfeeding surveillance systems may also enhance understanding of the real-world accessibility of lactation support. Additionally, expanding the analysis to include other types of lactation professionals and examining awareness of services among postpartum individuals could provide valuable insights. Longitudinal research tracking changes in IBCLC workforce density and breastfeeding rates over time would further clarify the nature of these relationships.

Conclusions

Given that national breastfeeding rates remain subpar, widespread promotion of breastfeeding as the sociocultural norm is a paramount initiative. The statistical results of this research collectively indicate that IBCLC density plays a valuable role in promoting positive breastfeeding outcomes, even when considering potential confounding factors. The positive correlations between IBCLC density and breastfeeding rates at various time points are supported by the simple linear and multiple regression analyses, highlighting the significance of lactation consultants in achieving breastfeeding success. However, these findings should be interpreted in light of certain limitations, including the use of aggregated state-level data and exclusion of non-IBCLC lactation professionals. Future research would benefit from more granular geographic data, such as breastfeeding rates at the county or zip code level and IBCLC distribution. Such data would enable more precise correlation and may uncover regional disparities masked in state-level analysis. One approach to improve data granularity could involve integrating provider workforce registries with public health datasets at the substate level or collaborating with hospitals and health departments to collect real-time lactation support utilization data. Additionally, exploring longitudinal outcomes and incorporating combination feeding practices could enhance the ecological validity of future studies. Many mothers are not aware that they can receive comprehensive health care from a lactation consultant to prevent and solve breastfeeding complications. Even with a fair population density of IBCLCs in an area, not having an awareness of their services poses a barrier to that care. Improved public awareness and policy-level initiatives that fund and integrate IBCLCs into postnatal care structures may significantly boost breastfeeding continuation rates. Further research with more granular baseline data would be an excellent start for promoting the expansion of IBCLC care and improving breastfeeding rates.

Acknowledgments

This study utilized secondary public datasets from the Centers for Disease Control and Prevention (CDC), the US Census Bureau, and the International Board of Lactation Consultant Examiners (IBLCE) and would not have been possible without their efforts to make this data accessible. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. 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.

In accordance with JMIR Publications' editorial policy on the use of generative AI tools, the authors affirm that no generative AI tools were used in the ideation, drafting, or editing of this manuscript. All content was written, reviewed, and fact-checked solely by the human authors. No text was generated or modified using AI assistance, and no AI-generated references were included.

Data Availability

All data used in this study are publicly available. International Board Certified Lactation Consultant workforce statistics were obtained from the International Board of Lactation Consultant Examiners, breastfeeding outcome data were sourced from the Centers for Disease Control and Prevention's 2022 Breastfeeding Report Card, and demographic data were obtained from the US Census Bureau. No individual-level or restricted-access datasets were used.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definition of terms.
[DOCX File, 15 KB - i-jmr_v14i1e70098_app1.docx]

XSL•FO
RenderX

Multimedia Appendix 2
Sequentially sorted data: International Board Certified Lactation Consultants density and breastfeeding rates.
[DOCX File, 18 KB - i-jmr_v14i1e70098_app2.docx]

References

1. Rollins N, Piwoz E, Baker P, et al. Marketing of commercial milk formula: a system to capture parents, communities, science, and policy. Lancet 2023 Feb 11;401(10375):486-502. [doi: 10.1016/S0140-6736(22)01931-6] [Medline: 36764314]
2. Stevens EE, Patrick TE, Pickler R. A history of infant feeding. J Perinat Educ 2009;18(2):32-39. [doi: 10.1624/105812409X426314] [Medline: 20190854]
3. Lactation consultant. Cleveland Clinic. 2021. URL: https://my.clevelandclinic.org/health/articles/22106-lactation-consultant [accessed 2024-02-17]
4. Breastfeeding report card. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/breastfeeding-data/media/pdfs/2024/06/2022-Breastfeeding-Report-Card-508.pdf [accessed 2024-02-17]
5. Perrine CG, Scanlon KS, Li R, Odom E, Grummer-Strawn LM. Baby-friendly hospital practices and meeting exclusive breastfeeding intention. Pediatrics 2012 Jul;130(1):54-60. [doi: 10.1542/peds.2011-3633] [Medline: 22665406]
6. Chrzan-Dętkoś M, Walczak-Kozłowska T, Pietkiewicz A, Żołnowska J. Improvement of the breastfeeding self-efficacy and postpartum mental health after lactation consultations - observational study. Midwifery 2021 Mar;94:102905. [doi: 10.1016/j.midw.2020.102905] [Medline: 33360178]
7. Patel S, Patel S. The effectiveness of lactation consultants and lactation counselors on breastfeeding outcomes. J Hum Lact 2016 Aug;32(3):530-541. [doi: 10.1177/0890334415618668] [Medline: 26644419]
8. Statistical report. IBLCE. URL: https://iblce.org/wp-content/uploads/2025/03/2025_March-24_IBCLCs_US_and_Territories_FINAL.pdf [accessed 2024-02-17]
9. Explore your state. US Census Bureau. URL: https://data.census.gov/profile/United_States?g=010XX00US#explore-your-state [accessed 2024-02-17]
10. State population by characteristics: 2020-2024. US Census Bureau. 2023 Jun 20. URL: https://www.census.gov/data/tables/time-series/demo/popest/2020s-state-detail.html [accessed 2024-02-17]
11. Indicators: metadata. World Health Organization. URL: https://platform.who.int/data/maternal-newborn-child-adolescent-ageing/indicator-explorer-new/MCA/women-of-reproductive-age-(15-49-years)-population-(thousands) [accessed 2024-02-17]
12. Braveman PA, Heck K, Egerter S, et al. The role of socioeconomic factors in Black-White disparities in preterm birth. Am J Public Health 2015 Apr;105(4):694-702. [doi: 10.2105/AJPH.2014.302008] [Medline: 25211759]

Abbreviations

CDC: Centers for Disease Control and Prevention
IBCLC: International Board Certified Lactation Consultants
IBLCE: International Board of Lactation Consultant Examiners
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology

Edited by TDA Cardoso; submitted 15.12.24; peer-reviewed by C Reischl, E Young, L Furman, S Harvey; revised version received 02.05.25; accepted 05.05.25; published 17.07.25.

Please cite as:
Dockins JF, Pahl HD, Lingerfelt DJ
Lactation Consultant Access and Breastfeeding Outcomes in the United States: Cross-Sectional Analysis
Interact J Med Res 2025;14:e70098
URL: https://www.i-jmr.org/2025/1/e70098
doi:10.2196/70098

© James F Dockins, Heather D Pahl, David J Lingerfelt. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 17.7.2025. 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 the COVID-19 Pandemic on Contraceptive Services at Selected Primary Health Care Facilities in India, Nigeria, and Tanzania: Cross-Sectional Study

Rita Kabra[1], MPH; Beena Joshi[2], PhD; Ester Elisaria[3], PhD; Tanimola Makanjuola Akande[4], MSc; Komal Preet Allagh[1], MPH; Adesola Olumide[5], PhD; Deepti Tandon[6], MS; Ranjan Prusty[7], PhD; Mary Ramesh[3], MPH; Donat Shamba[3], MSc; Bhavya MK[2], MSW; Shabana Khan[2], MPH; James Kiarie[1], MMed

[1]Department of Sexual and Reproductive Health and Research, including UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction, World Health Organization, Avenue Appia 20, Geneva, Switzerland

[2]Department of Operational and Implementation Research, ICMR National Institute for Research in Reproductive and Child Health, Mumbai, India

[3]Department of Health System Impact Evaluation and Policy, Ifakara Health Institute, Dar es Salaam, United Republic of Tanzania

[4]Department of Epidemiology and Community Health, University of Ilorin, Ilorin, Nigeria

[5]Institute of Child Health, University of Ibadan, Ibadan, Nigeria

[6]Department of Clinical Research, ICMR-National Institute for Research in Reproductive and Child Health, Mumbai, India

[7]Department of Biostatistics, ICMR- National Institute for Research in Reproductive and Child Health, Mumbai, India

Corresponding Author:
Rita Kabra, MPH
Department of Sexual and Reproductive Health and Research, including UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction, World Health Organization, Avenue Appia 20, Geneva, Switzerland

Abstract

Background: The COVID-19 pandemic disrupted sexual and reproductive health services, including family planning (FP) and contraceptive services. The World Health Organization conducted a multicountry study in India, Nigeria, and Tanzania to determine the impact of the pandemic on the health system's readiness to provide contraception services and trends in contraceptive uptake.

Objective: This study aimed to determine the status, availability, and health facility readiness to provide contraceptive services and to compare trends in contraceptive uptake before and during the pandemic.

Methods: This cross-sectional study was conducted by the Indian Council of Medical Research-National Institute of Research in Reproductive and Child Health (India), the University of Ilorin Teaching Hospital (Nigeria), and the Ifakara Health Institute (Tanzania). A total of 50 primary health facilities (11 in India, 6 in Nigeria, and 33 in Tanzania) were evaluated using a standardized facility assessment questionnaire, completed by the most knowledgeable senior health care provider or administrator at the facility. Monthly data on service utilization and contraceptive availability were collected to capture trends before and during the COVID-19 pandemic. Data were collected from May to August 2022. The study received ethical and scientific approval from the World Health Organization Ethics Review Committee and Research Project Review Panel and national regulatory bodies. Key outcomes included availability of FP guidelines and tools, service disruptions including contraceptive and abortion services, stock-outs, reasons for service disruptions, and mitigation measures to sustain service deliveries. Descriptive analysis was used to summarize the key trends and patterns.

Results: Health facilities in all three countries reported shortages of various contraceptives. Contraceptive services were partially disrupted in 91% facilities in India, 83% facilities in Nigeria, and 43% facilities in Tanzania. Abortion services were partially disrupted in all surveyed facilities offering these services in India and Nigeria and in 26.7% of facilities in Tanzania. Client visits declined in health facilities in 2020 compared to 2019 in India (30%) and Nigeria (11%), with a gradual recovery thereafter. In contrast, Tanzania experienced a 1% decline in client visits in 2020. Readiness measures such as telemedicine, task shifting, community outreach, triaging, and patient redirection were implemented to minimize service disruptions.

Conclusions: This study provides crucial insights into the challenges posed by the COVID-19 pandemic on contraceptive services and the measures taken to alleviate them. The findings can help countries to better prepare to prevent the disruption of FP and contraceptive services in future pandemics or emergencies.

(Interact J Med Res 2025;14:e59874) doi:10.2196/59874

XSL•FO
RenderX

KEYWORDS

facility assessment; COVID-19 pandemic; family planning; contraceptives; India; Nigeria; Tanzania; World Health Organization; contraceptive; COVID-19; primary health care; mixed method

Introduction

The COVID-19 outbreak was first reported in December 2019, rapidly spreading across the globe and being declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [1]. In an effort to curb the virus's transmission, countries implemented various measures such as international travel restrictions, social distancing, and lockdowns. The virus's highly dynamic nature posed challenges for taking standard coping measures in response to the pandemic. Every nation responded differently to the pandemic. Countries worldwide mobilized their resources to diagnose and treat COVID-19 patients while implementing public health strategies to contain the spread of the virus. Although necessary, this response could disrupt the provision of sexual and reproductive health (SRH) services, including family planning (FP) and contraceptive services, especially in low- and middle-income countries, where health systems are more vulnerable.

Several studies show that the COVID-19 pandemic had a major impact on the availability and provision of contraception, abortion, postabortion, and other sexual health services [2-4]. Supply chain disruptions limited the production, distribution, and availability of contraceptive commodities, resulting in stockouts [5]. Some health care facilities reduced or stopped offering their services [6], some health care providers were redirected to COVID-19 duties [7], and many women were unable to visit health care facilities due to lockdowns or fear of exposure to COVID-19 [8]. Travel to hospitals was disrupted by movement restrictions or reductions in health workers as a result of increasing numbers becoming infected with COVID-19 [9]. This created significant access and availability issues for many individuals who required these services. If the contraceptive needs of women and couples are unmet, unwanted pregnancies will increase, with long-term consequences for both women and their families [10]. It was estimated that significant levels of lockdown-related disruption over 6 months would leave 47 million women in low- and middle-income countries unable to use modern contraceptives, leading to a projected 7 million additional unintended pregnancies [11]. Furthermore, studies have modeled that even a 10% reduction in essential SRH services could lead to an estimated 15 million unintended pregnancies, 3.3 million unsafe abortions, and 29,000 additional maternal deaths in 12 months [10,12,13].

Our study aimed to determine the impact of the COVID-19 pandemic on the health system's capacity to provide contraception services in India, Nigeria, and Tanzania. These countries were selected because they were part of the World Health Organization's FP Accelerator Project [14], with expressed interest from the Ministries of Health to conduct the study. In this paper, we present findings of a health facility assessment, examining the availability of contraceptive services at primary health facilities and their readiness to provide these services during the COVID-19 pandemic. Additionally, we analyze and compare trends in contraceptive uptake in the 3 countries, both before and during the pandemic.

Methods

Study Objectives

The larger study used a mixed methods approach, using both quantitative and qualitative methods. The detailed methods and objectives of the overall study can be found in the published study protocol [15]. There were no deviations from the published protocol. In this paper, we present findings from the cross-sectional health facility assessment (quantitative study). This assessment was conducted in primary health facilities across India, Nigeria, and Tanzania. The primary objectives of the health facility assessment were to (1) determine the status, availability, and health facility readiness to provide contraceptive services in regions or localities most affected during the COVID-19 pandemic and (2) assess trends in the provision and uptake of contraceptive services in the facilities over the past 12 months. In this paper, we compare the findings of the health facility assessment from the 3 countries.

Study Design

This study was a cross-sectional health facility assessment using a structured questionnaire. Health care providers from each health facility were purposively selected to assist in completing the questionnaire. As part of the health facility assessment, we reviewed records related to contraception services to determine the availability, type, and range of contraceptive methods offered.

Settings

In India, this study was conducted by the Indian Council of Medical Research-National Institute of Research in Reproductive and Child Health, in Nigeria by the University of Ilorin Teaching Hospital, and in Tanzania by the Ifakara Health Institute. In the selected geographies, representative rural and urban primary health care (PHC) facilities with qualified staff providing contraception services were selected.

In India, Thane district in the State of Maharashtra was selected because it was one of the worst affected districts by the COVID-19 pandemic in Maharashtra [1]. Administratively, Thane district is divided into 7 talukas. A large part of this district is urban and is administratively governed by 6 municipal corporations. One taluka from an urban area (Navi Mumbai Municipal Corporation) and one from the rural area (Ambernath) were randomly selected [16]. In each taluka, all public health facilities were considered for the study.

In Nigeria, 3 zones (ie, South-West, North-Central, and North-West) of the 6 geopolitical zones were selected purposively. Within each zone, 1 state that is a site for Gates-funded SRH interventions (the Nigerian Urban Reproductive Health Initiative intervention and the Challenge Initiative) was selected for the study: Kwara (North-Central),

XSL•FO
RenderX

Oyo (South-West), and Kano (North West) states. Two high-volume PHC facilities that were providing maternal and child health and FP services were selected within each state (1 in an urban and 1 in a rural/periurban LGAs).

In Tanzania, 3 districts from 3 regions were selected: Temeke district (Dar es Salaam), Nyamagana district (Mwanza), and Moshi urban district (Kilimanjaro). Moshi urban and Temeke districts were selected as they were a part of the Service Availability and Readiness Assessment and belonged to regions with a high number of COVID-19 cases. The Nyamagana district was chosen as it was thought that COVID-19 would impact the urban population more because of their frequent social interactions. All government hospitals, health centers, and a random sample of dispensaries that offered FP services were selected in each district. Data collection was done between May 2022 and August 2022.

Participants

In each health facility, the most knowledgeable respondent about FP and contraceptive services was identified in consultation with the facility-in-charge. This individual was typically a senior health care provider or facility administrator. To be eligible, respondents had to meet the following inclusion criteria: (1) employed at the facility for at least 6 months before the onset of the COVID-19 pandemic, (2) directly involved in the provision of FP and contraceptive services, and (3) knowledgeable about the facility's service readiness and availability of FP and contraceptive services.

Assessments

A structured questionnaire was developed to assess the readiness of health facilities to provide FP and contraceptive services during the pandemic. The questionnaire was adapted from the WHO's SARA tool [17] and the health facility checklist developed by the WHO Department of Health Emergency [18]. Key domains assessed included policies and plans, service maintenance and referrals, facility infrastructure, human resources, and availability of commodities. The questionnaire was translated into local languages at each study site: Hindi and Marathi in India, Yoruba and Hausa in Nigeria, and Kiswahili in Tanzania.

The primary outcomes of the facility assessment included the proportion of health facilities with FP guidelines and tools available, proportion of facilities where regular health services were disrupted during the pandemic, proportion of facilities that maintained FP services as part of essential health services, stockouts of contraceptives during the pandemic, reported reasons for disruption in FP service delivery (eg, interruption of contraceptive services due to facility closures or staff reassignments and supply chain disruptions affecting the availability of contraceptives), and measures taken by countries to mitigate the impact of the pandemic on FP service provision.

Study Size

Health facilities in each country were purposively selected based on geographical location (eg, ease of access and areas with a high incidence of COVID-19 cases), organization of FP and contraceptive service delivery, and epidemic status, where the COVID-19 pandemic was likely to have significantly disrupted service delivery. A total of 50 health facilities were assessed across the 3 countries: 11 in India, 6 in Nigeria, and 33 in Tanzania. Of the 33 health facilities in Tanzania, 3 were government hospitals, 9 were health centers, and 21 were dispensaries. In Nigeria, 4 PHC facilities, 1 district health unit, and 1 maternal and child health facility were included. Of the 11 facilities in India, 5 were urban health posts, 4 were PHCs, 1 was district hospital, and 1 was rural hospital. Table 1 presents the details of the study sites in each country.

XSL•FO
RenderX

Table . Description of study sites in India, Nigeria, and Tanzania.

Country and location	Setting
India (n=11)	
Navi Mumbai Municipal Corporation area, Thane district (urban)	5 urban health posts (PHCs[a] in urban areas) and 1 district hospital
Ambernath block (rural)	4 PHCs and 1 rural hospital
Nigeria (n=6)	
Kwara state, Ilorin LGA[b]	1 PHC (urban)
Kwara state, Afon LGA	1 district health unit (rural)
Oyo state, Ibadan north	1 PHC (urban)
Oyo state, Akinyele	1 PHC (periurban)
Kano state, Dala LGA	1 MCH[c] (urban)
Kano state, Kura LGA	1 PHC (rural)
Tanzania (n=33)	
Mwanza, Nyamagana district	6 dispensaries, 2 health centers, and 1 hospital
Kilimanjaro, Moshi urban district	7 dispensaries, 2 health centers, and 1 hospital
Dar es Salaam, Temeke district	8 dispensaries, 5 health centers, and 1 hospital

[a]PHC: primary health care.

[b]LGA: local government areas.

[c]MCH: maternal and child health.

Data Analysis

In India, data from facility assessments were entered into an electronic database and analyzed using SPSS (Statistical Package for Social Sciences). In Nigeria, the KoboCollect app was used on the research staff's phones to collect data, and it was analyzed using SPSS version 21, whereas in Tanzania, data were collected digitally on tablets using the Open Data Kit platform and analyzed using STATA version 15 software. Descriptive analysis was used to compare the characteristics of the various facilities, including the number of clients, the type of services provided, and the contraceptive stocks.

Ethical Considerations

The WHO Research Project Review Panel granted scientific approval, whereas the WHO Ethics Review Committee granted ethical approval (protocol IDs CERC.0103K, CERC.0103J, and CERC.0103I). In India, the Institutional Ethical Committee of Indian Council of Medical Research-National Institute of Research in Reproductive and Child Health, Mumbai (D/ICEC/Sci-166/175/2021) approved the study protocol. Additionally, permission was obtained from state and district government health officials to conduct the study at the selected sites. In Nigeria, the study was approved by the National Health Research Ethics Committee of Nigeria (NHREC/01/01/2007-07/09/2021), the University of Ilorin Ethical Research Committee (in Kwara state), and State Ministries of Health Ethical Committee (in Kwara, Kano, and Oyo States). In Tanzania, the study was approved by the Ifakara Health Institutional Review Board (IHI/IRB/NO. 44 - 2021) and the National Institute of Medical Research (NIMR/HQ/R.8a/Vol.IX/3916). In addition, the team received permission letters to visit the selected sites from the President's Office, Regional and Local Government.

All participants were informed about the study and signed a written consent before participation. Data were deidentified using unique ID numbers to maintain confidentiality. Participation was entirely voluntary, and no financial incentives were provided. Patients or the public were not involved in the design or conduct or reporting or dissemination plans of this research.

Results

Health Facility Assessment

We compare and summarize the health facility assessment findings from the 3 countries in Table 2 and present a heat map of these findings in Multimedia Appendix 1.

XSL•FO
RenderX

Table . Study findings from India, Nigeria, and Tanzania.

	Tanzania	Nigeria	India
Availability of policies and funds for essential health services			
Defined national SRH essential service package before COVID-19 pandemic	93.9 %	100 %	100 %
Identified core set of essential health services to be maintained during COVID-19 pandemic	63.6 %	83.3 %	100 %
Received additional government funding to assure essential health services	63.6 %	83.3 %	100 %
Received WHO[a] guidelines on COVID-19 response and continuity of SRH[b] essential health services	54.4 %	83.3 %	100 %
Maintenance of essential health services			
Outpatient department services	91 %	16.7 %	100 %
Inpatient services	57.6 %	16.7 %	54.5 %
Emergency health services	93.9 %	16.7 %	100 %
Prehospital emergency care services	93.9 %	0%	91 %
Community-based care	91 %	16.7 %	36.4 %
Mobile clinics	72.7 %	0%	Not applicable
Maintenance of essential SRH services as normal (no disruption)			
Contraception services	54.5 %	16.7 %	9.1 %[c]
Abortion services	66.7 %	0%	0%[d]
ANC[e] services	66.7 %	33.3 %	63.6 %
Facility-based births	80 %	33.3 %	100 %
Routine immunization services	72.7 %	33.3 %	45.5 %
Sick child services/IMNCI[f]	75.8 %	33.3 %	81.8 %
Outbreak detection and control (for non-COVID diseases)	75.8 %	33.3 %	72.7 %
Inpatient critical care services	66.7 %	Not applicable	100 %
24 h emergency room	84 %	33.3 %	87.5 %
Availability of national FP[g] and abortion guidelines/job aids and service referrals			
Availaility of national FP guidelines in facility	100 %	83.3 %	100 %
Availability of FP checklists or job aids in facility	100 %	100 %	100 %
Cilents referred for FP services to other health care facilities	48.5 %	0%	91.8 %
Availability of national abortion guidelines in facility	51.5 %	16.7 %	100 %
Availability of safe abortion checklists or job aids in facility	39.4 %	0%	100 %

XSL•FO
RenderX

	Tanzania	Nigeria	India
Women who came to facility for postabortion care received FP counseling and services	57.6 %	50 %	69 %
Availability of contraceptive stocks (no stock out in the last 24 months)			
Combined oral contraceptive pill	48.5 %	16.7 %	90.9 %
Progestin-only contraceptive pill	46.9 %	33.3 %	Not applicable
Male condom	51.5 %	33.3 %	100 %
Female condom	43.8 %	16.7 %	Not applicable
Combined injectable contraceptives	63.3 %	16.7 %	Not applicable
Progestin-only injection	50 %	16.7 %	45.6 %
Centchroman	Not applicable	Not applicable	63.4 %
Cycle beads for standard days method	42.4 %	33.3 %	Not applicable
Vaginal ring	66.7 %	66.7 %	Not applicable
Subdermal implant	45.5 %	33.3 %	Not applicable
Copper IUD[h]	53.1 %	50 %	Not applicable
Levonorgestrel IUD	71.4 %	66.7 %	Not applicable
Diaphragm	78.9 %	33.3 %	Not applicable
Emergency contraceptive pill	41.9 %	50 %	100 %

[a]WHO: World Health Organization.

[b]SRH: sexual and reproductive health.

[c]Mainly due to the suspension of female sterilization services. Other contraceptive services were partially disrupted, leading to the upscaling of the community-based distribution of contraceptive commodities.

[d]Abortion services were available at only 2 of 11 facilities surveyed, and at both facilities, abortion services were partially disrupted.

[e]ANC: antenatal care.

[f]IMNCI: Integrated Management of Newborn and Childhood Illness

[g]FP: family planning.

[h]IUD: intrauterine device.

Availability of Policies and Funds for Essential Health Services

Before the COVID-19 pandemic, all facilities in Nigeria and India had policies on the national SRH essential health services package, compared to 94% of facilities in Tanzania (Table 2). All health facilities in India, 83% in Nigeria, and 64% in Tanzania identified a core set of essential health services to be maintained during the pandemic. Additional government funds were allocated to all facilities in India, 64% of facilities in Tanzania, and 83% in Nigeria to ensure the continuation of essential health services. Approximately half of Tanzania's facilities (54%) received the latest WHO guidelines on COVID-19 response and continuity of SRH services, compared to 100% in India and 83% in Nigeria. Overall, India demonstrated the highest preparedness and support for maintaining essential SRH services during the COVID-19 pandemic, with full policy coverage, guideline dissemination, and funding, whereas Tanzania lagged behind in all areas.

Maintenance of Essential Health Services

During the COVID-19 pandemic, the outpatient department (OPD) services functioned normally in all Indian facilities, 91% in Tanzania, and only 17% in Nigeria (Table 2). The OPD services were suspended in one facility in Tanzania. Inpatient services continued to operate as usual in 55% of facilities in India, 58% in Tanzania, and 17% in Nigeria, with suspensions reported in 1 facility each in India and Nigeria. Emergency health services had minimal disruptions in India and Tanzania but major disruptions in Nigeria (only 17% functioned as normal). Community-based care continued normally in most facilities (91%) in Tanzania, 1 in Nigeria (17%), and 4 (37%) in India. Most mobile clinics functioned normally in 73% of Tanzania's facilities, but none in Nigeria. In India, none of the included facilities offered mobile clinics. Thus, all essential health services were majorly disrupted in Nigeria, whereas India and Tanzania managed to maintain most services, with moderate disruptions in inpatient services.

Maintenance of Essential SRH Services

Contraceptive services continued as normal in only 1 facility each in India (9.1%) and Nigeria (16.7%) and 18 facilities

XSL•FO
RenderX

(54.5%) in Tanzania (Table 2). These services were partially disrupted by the pandemic in all 3 countries: 91% in India (mainly facility-based services such as sterilization services, which is the most common contraceptive method accepted overall in India); 83% in Nigeria; and 43% in Tanzania. One facility (3%) in Tanzania reported a complete disruption of contraceptive services, whereas no facility in India or Nigeria faced complete disruption of services. In fact, in India, outreach services were initiated through ASHAs for household distribution of condoms and OC pills. Abortion services were partially disrupted in both the facilities providing these services in India and Nigeria, and in 4 of the 15 (26.7%) facilities providing abortion services in Tanzania.

Antenatal care services were partially disrupted in 36% of facilities in India, 50% of facilities in Nigeria, and 33% of facilities in Tanzania. Facility-based births were not disrupted in India but experienced disruptions in 67% and 20% of facilities in Nigeria and Tanzania, respectively. Routine immunization services were partially disrupted in 46% of facilities in India, 67% in Nigeria, and 24% in Tanzania. One facility in India and Tanzania reported a complete disruption of immunization services. Inpatient critical care services were not disrupted in the Indian facilities, whereas they were partially disrupted in 33% of Tanzanian facilities. None of the included facilities in Nigeria offered inpatient critical care. In Nigeria, Tanzania, and India, emergency room services were partially disrupted in 67%, 16%, and 13% of the facilities.

Overall, Tanzania was the most successful among the 3 countries in maintaining essential SRH services during the pandemic, particularly contraceptive services, whereas India and Nigeria experienced partial disruptions, particularly in facility-based and abortion services. Nigeria was the most severely affected in maternal and child health services. Table 2 depicts the percentage of SRH services not disrupted during the pandemic.

Availability of National FP and Abortion Guidelines/Job Aids and Service Referrals

National FP guidelines, checklists, job aids, and contraceptive stocks were available in all facilities in the 3 countries except for FP national guidelines in 1 Nigerian facility (Table 2). Referrals for FP services were made in 92% of Indian facilities (especially for female sterilization) and 49% of Tanzanian facilities, but none in Nigeria. National abortion guidelines and safe abortion checklists/job aids were available at all facilities in India, whereas in Nigeria, national abortion guidelines were available in one facility (17%), and none had safe abortion checklists/job aids. Tanzania had 52% of facilities with the national abortion guidelines, and 39% with the safe abortion checklists/job aids. FP counseling and services for postabortion care were offered in 50% and 58% of facilities in Nigeria and Tanzania, respectively. In India, 69% (majorly in urban areas) of women received sterilization and IUCD services after an abortion at a health facility. Overall, India had the most robust infrastructure supporting FP and abortion services, including widespread availability of national guidelines, job aids, and referral systems. Nigeria exhibited significant gaps, particularly in abortion guidelines, checklists, and the absence of FP referrals during the pandemic, whereas Tanzania had moderate availability of national guidelines and job aids but limited referral services.

Availability of Contraceptive Stocks at Health Facilities

All 3 countries experienced contraceptive shortages during the pandemic (Table 2). Nigeria faced shortages of various contraceptives, including combined oral contraceptives (83%) and progestin-only pills (67%), injectables (both combined injectable contraceptives and progesterone only; 83%), male (67%) and female condoms (83%), and subdermal implants (67%). Approximately half of the facilities in Tanzania reported a shortage of both combined and progesterone-only pills, female condoms (56%), implants (55%), and emergency contraceptive pills (58%). India reported shortages of progestin-only injectables (55%) and centchroman (36%). Table 2 depicts the percentage of health facilities with no contraceptive stock out. Overall, Nigeria was the most severely impacted, with over two-thirds of facilities reporting shortages of multiple contraceptive methods. Tanzania experienced moderate contraceptive stockouts, whereas India reported fewer disruptions.

Trends in Clients Seeking FP and Contraceptive Services

Data from health facilities were used to track trends of modern contraceptive uptake before and during the COVID-19 pandemic (Figure 1 and Table 3). India saw a 30% drop in clients visiting health facilities for contraceptive services during Q2 2020 (first wave of the pandemic; Table 3), then a recovery with a 35% increase in Q2 2021 compared to Q2 2020. Tanzania showed minimal change with a slight dip (1%) in Q3 2020, whereas Nigeria experienced an 11% decrease in clients in Q4 2020, with a recovery in Q1 2021, followed by a second dip in Q2 2021 (17%). Overall, India demonstrated a strong recovery following an initial sharp decline. Tanzania maintained consistent service use with minimal fluctuation, whereas Nigeria experienced more prolonged disruptions.

XSL•FO
RenderX

Figure 1. Trends in contraceptive uptake at the included health facilities in India, Tanzania, and Nigeria.

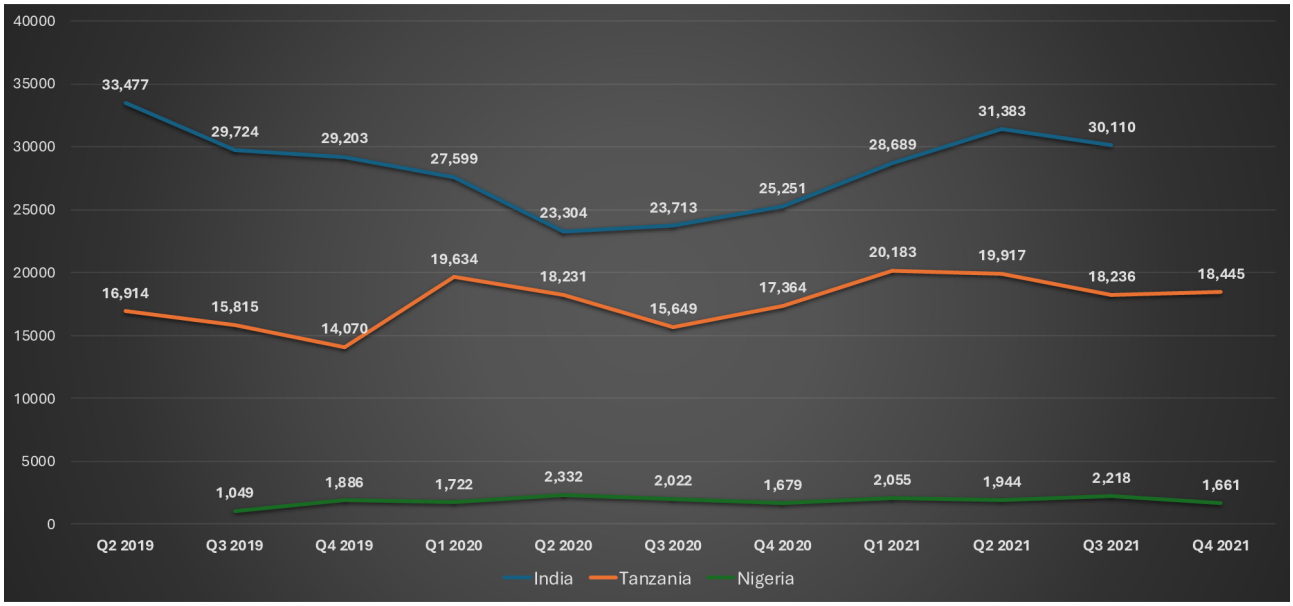


Table . Year-over-year percentage change in the contraceptive uptake at primary health facilities.

	Q1 2020 (%)	Q2 2020 (%)	Q3 2020 (%)	Q4 2020 (%)	Q1 2021 (%)	Q2 2021 (%)	Q3 2021 (%)	Q4 2021 (%)
India	[a]NA	−30.4	−20.2	−13.5	3.9	34.7	30	NA
Tanzania	20.1	7.8	−1	23.4	2.8	9.2	16.5	6.2
Nigeria	NA	NA	10.4	−11	19.3	−16.6	9.7	1

[a]Not applicable.

Main Causes of Disruption or Change in Service Utilization

The causes of service disruption varied among the 3 countries (Table 4). In Nigeria, essential health services were mainly disrupted due to insufficient staffing at health facilities (83%), staff being assigned to COVID-19 services (83%), lack of public transportation (83%), and decreased number of clients at OPD (83%). India experienced disruptions due to a decrease in outpatient volume (54.5%) and lack of public transportation (45.5%). In Tanzania, service disruption was attributed to decreased outpatient visits (68.2%), insufficient staff (68.2%), and inadequate personal protective equipment (PPE) kits (63.6%). Overall, Nigeria experienced severe disruptions driven by workforce and access-driven constraints. India faced moderate challenges due to reduced outpatient volumes and mobility issues, whereas in Tanzania, a combination of decreased demand, staffing limitations, and insufficient PPE were predominant factors. Table 4 compares the major reasons for disruption or change in service usage during the pandemic.

XSL•FO
RenderX

Table . Reasons for disruption/ change in service utilization in India, Nigeria, and Tanzania.

Major reasons for the change in health service utilization	India (%)	Nigeria (%)	Tanzania (%)
Closure of OPD[a] services	0	33.3	4.6
Closure of disease-specific consultation clinics	0	33.3	4.6
Decrease in OPD volume due to patients not presenting	54.5	83.3	68.2
Decrease in in-patient volume due to cancelation of elective care	18.2	50	22.7
Inpatient services are not available	18.2	0	0
Insufficient staff to provide services	36.4	83.3	68.2
Related clinical staff deployed to provide COVID-19 relief	36.4	83.3	50
Insufficient PPE available for healthcare providers	9.1	50	63.6
Stock out of essential medicines, medical diagnostics, or other health products	18.2	66.7	45.5
Changes in treatment policies for care-seeking behavior for fever symptoms	9.1	66.7	31.8
Government transport lockdowns hindering access to health facilities	45.5	83.3	18.2
Financial difficulties during outbreak/lockdown	18.2	66.7	27.3

[a]OPD: outpatient department.

Approaches to Mitigate/Overcome the Disruptions to Essential Health Services

Countries adopted various measures to ensure the continuity of essential health services (Table 5). In India, Nigeria, and Tanzania, 91%, 83%, and 42% of facilities, respectively, implemented task shifting. Telemedicine services were provided by 82% of facilities in India, 50% in Nigeria, and 24% in Tanzania. When services were unavailable, patients were redirected to alternate health care facilities by all facilities in India, 67% of facilities in Nigeria, and 42% of facilities in Tanzania. In Nigeria, 83% of facilities informed the community about service disruptions and changes, whereas this was done by 42% and 18% of facilities in Tanzania and India. All facilities in India and 67% and 55% of facilities in Nigeria and Tanzania triaged clients to identify priority. In Tanzania, Nigeria, and India, user fees were removed in 6%, 17%, and 46% of facilities, respectively.

Table . Approaches used by countries to overcome disruption to essential health services during the pandemic.

Approach	India (%)	Nigeria (%)	Tanzania (%)
Telemedicine	81.8	50	24.2
Task shifting	90.9	83.3	42.4
Novel supply chain for medicines	54.5	33.3	3
Triaging to identify priorities	100	66.7	54.6
Redirection of patients to alternate health facilities	100	66.7	42.4
Community outreach to inform on service disruptions and changes	18.2	83.3	42.4
Government removal of user fees	45.5	16.7	6.1

XSL•FO
RenderX

Discussion

Principal Findings and Comparative Evidence

We determined the availability of contraceptive services in selected primary health facilities during the COVID-19 pandemic and compared the trends in contraceptive uptake before and during the pandemic in India, Nigeria, and Tanzania. Our study revealed that the COVID-19 pandemic resulted in partial disruptions in contraceptive services across the majority of health facilities in India and Nigeria and less than half of the facilities in Tanzania, particularly during the first wave of the pandemic. Notably, 1 facility in Tanzania even reported a complete disruption in contraceptive services. According to the WHO Pulse Survey, conducted in 2020 and 2021, 59% and 44% of countries, respectively, experienced partial disruptions, whereas 9% and 5% reported severe disruptions in FP and contraceptive services [19,20]. A study from South Asia showed a 3% to 31% decline in FP coverage during the period of stringent COVID-19 restrictions in 2020 [21]. Another study from sub-Saharan Africa showed a reduction in FP consultations by 10% to 25% between March and July 2020 in 4 of the 7 countries [22]. Our study noted a decline of 17% and 30% clients seeking contraceptive services in Nigeria and India during the initial wave of the pandemic in 2020, with a subsequent recovery in the number of clients. Literature suggests that despite initial declines during the pandemic, many countries were able to maintain or restore access to essential health services, including FP [13]. A multicountry cross-sectional study (Nepal, Uganda, Niger, and Malawi) reported that the pandemic reduced the number of long-acting reversible contraceptive users from 5% in Nepal to 22% in Malawi [23]. Contrary to the disruptions observed in India and Nigeria, our study showed a consistent increase in the usage of contraceptive services in Tanzania from 2019 to 2021. This upward trend could be attributed to Tanzania's decision to refrain from imposing strict lockdown measures, allowing for unrestricted movement, which enabled them to maintain their FP clients. Similar results were reported from another study from Tanzania, which showed a slight dip in FP usage among existing users in 2020, followed by a rise in the number of new contraceptive users during 2020 and 2021 compared to the same period in 2019 [24]. All 3 countries experienced stock-outs of various contraceptives during the pandemic, affecting 36% to 83% of facilities. The United Nations Population Fund predicted in 2020 that most countries would experience stockouts of 1 or more modern contraceptive methods [8].

Our research identified several factors contributing to the disruption of contraceptive services during the COVID-19 pandemic. Main causes included a decrease in women visiting the facility, lack of public transportation, insufficient staff at health facilities, and staff being reassigned to COVID-19–related duties. These findings align with other studies that highlighted how the pandemic impacted women's access to contraception due to supply chain disruptions, health providers being reassigned to COVID-related duties, and transportation issues caused by lockdowns [20,25-28]. According to the 2020 WHO Pulse survey, the main reasons for disruption in health services were patients not presenting to outpatient care (76%), transport lockdowns hindering access (48%), and financial difficulties during the pandemic (33%). On the supply side, the common causes were cancelation of elective care (66%), staff deployed to provide COVID-19 relief (49%), and insufficient PPE equipment for health providers (44%) [19]. A scoping review on barriers to accessing contraceptive services during the pandemic included fear of infection, lack of transport, cost, longer waiting times, reduced home visits by health workers, and limited availability of contraceptive supplies as the main barriers [2]. Findings from another study in Tanzania indicate factors for disruption in contraceptive services as changes in clinic schedules, fear of COVID-19 infection, and a shortage of contraceptives [24].

Countries implemented different approaches to overcome disruptions owing to the pandemic, such as redirecting clients to alternative health facilities, triaging to identify priorities, task shifting to lower-level health workers, using telemedicine, and conducting community outreach to inform clients about service disruptions. These strategies align with those reported in the WHO Pulse Survey, where triaging was the most common response (76%), followed by telemedicine (63%) and task shifting (57%) [19].

While our study offers valuable insights into how the pandemic affected contraceptive services and the strategies used by countries to address the crisis, it is essential to recognize that the study's geographic coverage is limited, and the number of health care facilities examined was relatively small. These limitations could affect the generalizability of the study's findings. Nonetheless, the study offers crucial insights into the challenges posed by the pandemic and the measures taken to mitigate them.

Conclusions

Our study highlighted varying degrees of disruptions in essential health services and SRH services across the 3 countries. The most significant disruptions were observed in Nigeria, whereas Tanzania experienced comparatively fewer disruptions. Common factors contributing to service disruption in all 3 countries included reduced OPD visits and the lack of public transportation due to the lockdown. Mitigation strategies such as telemedicine, task shifting, triaging, community outreach, and redirection of patients proved effective in alleviating these disruptions. Moving forward, it is crucial for countries to implement and standardize these strategies to be better prepared for future pandemics.

XSL•FO
RenderX

Acknowledgments

The authors would like to express their gratitude to staff in the World Health Organization Department of Sexual and Reproductive Health and Research that includes the UNDP-UNFPA-UNICEF-WHO-World Bank Special Programme of Research, Development and Research Training in Human Reproduction. Dr. Moazzam Ali and Dr. Caron Rahn Kim developed the generic protocol.

Funding

This research is funded by the Canadian grant to the WHO Department of Sexual and Reproductive Health and Research, which includes the United Nations Development Programme, the United Nations Population Fund, United Nations International Children's Emergency Fund, World Health Organization, World Bank Special Programme of Research, and Development and Research Training in Human Reproduction (Human Reproduction Programme). The funding agency has no role in designing, implementing, analyzing, and reporting the study results.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JK, RK, BJ, EE, TMA
Formal analysis: KPA, RK, AO, DT, RP, MR, DS, BMK, SK
Writing – original draft: KPA, RK, JK
Writing – review and editing: all authors

Conflicts of Interest

None declared.

Multimedia Appendix 1
Heatmap comparing study findings from India, Nigeria, and Tanzania
[PDF File, 219 KB - i-jmr_v14i1e59874_app1.pdf]

References

1. Timeline: WHO's COVID-19 response world health organization. World Health Organization. URL: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline [accessed 2023-06-08]
2. Polis CB, Biddlecom A, Singh S, Ushie BA, Rosman L, Saad A. Impacts of COVID-19 on contraceptive and abortion services in low- and middle-income countries: a scoping review. Sex Reprod Health Matters 2022 Dec;30(1):2098557. [doi: 10.1080/26410397.2022.2098557] [Medline: 35920612]
3. VanBenschoten H, Kuganantham H, Larsson EC, et al. Impact of the COVID-19 pandemic on access to and utilisation of services for sexual and reproductive health: a scoping review. BMJ Glob Health 2022 Oct;7(10):e009594. [doi: 10.1136/bmjgh-2022-009594] [Medline: 36202429]
4. Mukherjee TI, Khan AG, Dasgupta A, Samari G. Reproductive justice in the time of COVID-19: a systematic review of the indirect impacts of COVID-19 on sexual and reproductive health. Reprod Health 2021 Dec 20;18(1):252. [doi: 10.1186/s12978-021-01286-6] [Medline: 34930318]
5. Chris P. Opinion: how will COVID-19 affect global access to contraceptives and what can we do about? devex. 2020. URL: https://www.devex.com/news/opinion-how-will-covid-19-affect-global-access-to-contraceptives-and-what-can-we-do-about-it-96745 [accessed 2023-09-14]
6. Stories from the frontline: in the shadow of the COVID-19 pandemic. MSI. 2020. URL: https://www.msichoices.org/covid-19/stories-from-the-frontline/ [accessed 2023-09-14]
7. Sarita S. Family planning methods upended by the coronavirus. FP. 2020. URL: https://foreignpolicy.com/2020/05/13/india-family-planning-upended-coronavirus-women-sexual-reproductive-health/ [accessed 2023-09-14]
8. Impact of the COVID-19 pandemic on family planning and ending gender-based violence, female genital mutilation and child marriage. UNFPA. 2020. URL: https://www.unfpa.org/resources/impact-covid-19-pandemic-family-planning-and-ending-gender-based-violence-female-genital [accessed 2023-09-14]
9. Riley T, Sully E, Ahmed Z, Biddlecom A. Estimates of the potential impact of the COVID-19 pandemic on sexual and reproductive health in low- and middle-income countries. Int Perspect Sex Reprod Health 2020 Apr 16;46(May):73-76. [doi: 10.1363/46e9020] [Medline: 32343244]
10. Dasgupta A, Kantorová V, Ueffing P. The impact of the COVID-19 crisis on meeting needs for family planning: a global scenario by contraceptive methods used [version 2; peer review: 1 approved, 2 approved with reservations]. Gates Open Research. 2020. URL: https://gatesopenresearch.org/articles/4-102/v2 [accessed 2023-09-14]

XSL•FO
RenderX

11. Millions more cases of violence, child marriage, female genital mutilation, unintended pregnancy expected due to the COVID-19 pandemic. United Nations Population Fund. URL: https://www.unfpa.org/news/millions-more-cases-violence-child-marriage-female-genital-mutilation-unintended-pregnancies [accessed 2023-09-14]
12. Coronavirus disease (COVID-19) pandemic: UNFPA response plan. UNFPA. 2020. URL: https://www.unfpa.org/resources/coronavirus-disease-covid-19-pandemic-unfpa-global-response-plan [accessed 2023-09-14]
13. Impact of COVID-19 on family planning: what we know one year into the pandemic. UNFPA. URL: https://www.unfpa.org/resources/impact-covid-19-family-planning-what-we-know-one-year-pandemic [accessed 2023-09-14]
14. The WHO family planning accelerator project. World Health Organization. URL: https://www.who.int/teams/sexual-and-reproductive-health-and-research-(srh)/areas-of-work/family-planning-and-contraception/access-to-quality-family-planning-services [accessed 2023-06-08]
15. Kabra R, Joshi B, Elisaria E, et al. Determining the impact of the COVID-19 pandemic on availability, use and readiness of family planning and contraceptive services at selected primary health care facilities in Africa and Asia: protocol for a mixed methods study. JMIR Res Protoc 2023 May 10;12:e43329. [doi: 10.2196/43329] [Medline: 36927830]
16. About district. Government of Maharashtra District Thane. URL: https://thane.nic.in/about-district/ [accessed 2023-09-14]
17. Service availability and readiness assessment (SARA). World Health Organization. URL: https://www.who.int/data/data-collection-tools/service-availability-and-readiness-assessment-(sara)?ua=1#:~:text=The%20Service%20Availability%20and%20Readiness,managing%20of%20a%20health%20system [accessed 2023-09-14]
18. Hospital emergency response checklist. World Health Organization. URL: https://www.who.int/publications/i/item/hospital-emergency-response-checklist [accessed 2023-09-14]
19. Second round of the national pulse survey on continuity of essential health services during the COVID-19 pandemic: january-march 2021. World Health Organization. 2021 Apr 22. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-EHS-continuity-survey-2021.1 [accessed 2023-09-14]
20. Pulse survey on continuity of essential health services during the COVID-19 pandemic. World Health Organization. 2020 Aug. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-EHS_continuity-survey-2020.1 [accessed 2023-09-12]
21. Direct and indirect effects of the COVID-19 pandemic and response in south asia. UNICEF. 2021. URL: https://www.unicef.org/rosa/media/13066/file/Main%20Report.pdf [accessed 2023-09-14]
22. Shapira G, Ahmed T, Drouard SHP, et al. Disruptions in maternal and child health service utilization during COVID-19: analysis from eight sub-Saharan African countries. Health Policy Plan 2021 Aug 12;36(7):1140-1151. [doi: 10.1093/heapol/czab064] [Medline: 34146394]
23. Brunie A, Austin G, Arkin J, Archie S, Amongin D, Ndejjo R, et al. Women's experiences with family planning under COVID-19: a cross-sectional interactive voice response survey in malawi, nepal, niger and uganda. Glob Health Sci Pract. 2022 Aug. URL: https://www.ghspjournal.org/content/10/4/e2200063 [accessed 2023-09-06]
24. Shayo EH, Nassor NK, Mboera LEG, et al. The impacts of COVID-19 and its policy response on access and utilization of maternal and child health services in Tanzania: a mixed methods study. PLOS Glob Public Health 2023;3(5):e0001549. [doi: 10.1371/journal.pgph.0001549] [Medline: 37172038]
25. Roy N, Amin MB, Maliha MJ, et al. Prevalence and factors associated with family planning during COVID-19 pandemic in Bangladesh: a cross-sectional study. PLoS ONE 2021;16(9):e0257634. [doi: 10.1371/journal.pone.0257634] [Medline: 34547041]
26. Ullah MA, Moin AT, Araf Y, Bhuiyan AR, Griffiths MD, Gozal D. Potential effects of the COVID-19 pandemic on future birth rate. Front Public Health 2020;8:578438. [doi: 10.3389/fpubh.2020.578438] [Medline: 33363080]
27. Otu Michael T, Agbana RD, Ojo TF, Kukoyi OB, Ekpenyong AS, Ukwandu D. COVID-19 pandemic and unmet need for family planning in Nigeria. Pan Afr Med J 2021;40:186. [doi: 10.11604/pamj.2021.40.186.27656]
28. Aly J, Haeger KO, Christy AY, Johnson AM. Contraception access during the COVID-19 pandemic. Contracept Reprod Med 2020;5:17. [doi: 10.1186/s40834-020-00114-9] [Medline: 33042573]

Abbreviations

FP: family planning
PHC: primary health care
PPE: personal protective equipment
SRH: sexual and reproductive health
WHO: World Health Organization

XSL•FO
RenderX

Edited by N Cahill, TDA Cardoso; submitted 25.04.24; peer-reviewed by K Zhang; revised version received 19.08.25; accepted 20.08.25; published 02.12.25.

Please cite as:
Kabra R, Joshi B, Elisaria E, Akande TM, Allagh KP, Olumide A, Tandon D, Prusty R, Ramesh M, Shamba D, MK B, Khan S, Kiarie J
Impact of the COVID-19 Pandemic on Contraceptive Services at Selected Primary Health Care Facilities in India, Nigeria, and Tanzania: Cross-Sectional Study
Interact J Med Res 2025;14:e59874
URL: https://www.i-jmr.org/2025/1/e59874
doi:10.2196/59874

© Rita Kabra, Beena Joshi, Ester Elisaria, Tanimola Makanjuola Akande, Komal Preet Allagh, Adesola Olumide, Deepti Tandon, Ranjan Prusty, Mary Ramesh, Donat Shamba, Bhavya MK, Shabana Khan, James Kiarie. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 2.12.2025. 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

Student and Physician Views of How the Dobbs Decision Affects Training and Practice Location Preferences: Cross-Sectional Questionnaire Study

Morgan S Levy[1*], MPH, MD; Simone A Bernstein[2*], MHPE, MD; Sarah M McNeilly[3], BA; Abigail Liberty[4], MSPH, MD; Shira Fishbach[5], MD; Shikha Jain[6], MD; Jessica A Gold[2], MS, MD; Vineet M Arora[7], MAPP, MD

[1]Department of Radiation Oncology, University of Kentucky College of Medicine, Lexington, KY, United States

[2]Department of Psychiatry, Washington University School of Medicine in St. Louis, St. Louis, MO, United States

[3]Albert Einstein College of Medicine, New York, NY, United States

[4]Department of Obstetrics and Gynecology, Oregon Health and Science University, Portland, OR, United States

[5]Department of Obstetrics and Gynecology, University of Michigan School of Medicine, Ann Arbor, MI, United States

[6]Department of Medicine, University of Illinois at Chicago, Chicago, IL, United States

[7]Department of Medicine, University of Chicago Pritzker School of Medicine, Chicago, IL, United States

[*]these authors contributed equally

Corresponding Author:
Morgan S Levy, MPH, MD
Department of Radiation Oncology
University of Kentucky College of Medicine
800 Rose Street
Lexington, KY, 40536
United States
Phone: 1 9733092423
Email: morgan.levy4@gmail.com

Abstract

Background: By allowing for abortion bans and restrictions to take effect in the majority of US states, the 2022 *Dobbs v Jackson Women's Health Organization* decision portends to have lasting impacts on patient care and the physician workforce. Notably, it is already beginning to impact practice location preferences of US health care workers, evidenced by declining application rates to residency programs in abortion-restrictive states since 2022. Yet, there remains a gap in the literature regarding why this trend exists.

Objective: This study aims to describe what factors are driving the practice location preferences of medical students and physicians after the *Dobbs* decision.

Methods: This study analyzes qualitative data from a web-based, cross-sectional study. In August 2022, a nonprobabilistic sample of physicians and medical students were surveyed on social media about the impact of overturning *Roe v Wade* on practice location preferences, which included the free-text question "Please share your thoughts about the overturning of *Roe v Wade* and how it will affect your decision about your (residency/job or fellowship) programs." A total of 3 independent team members completed an inductive thematic analysis of 524 free responses, resolving differences by discussion.

Results: Approximately 1 in 4 survey respondents also completed the free-response item (524/2063, 25.4%); a total of 219 were medical students, 129 were residents and fellows, and 176 were practicing physicians. Of them, approximately half (261/524, 50.5%) resided in states where abortion bans were in place or anticipated. Those who answered the free-response item were relatively more likely to hail from states with restrictive abortion bans (P<.001) compared to those who did not, with other demographic characteristics being largely similar between the groups. Inductive thematic analysis yielded 2 broad thematic categories: patient-related and workforce-related factors influencing practice decision preferences. The 3 most common themes overall were respondent concerns regarding their patient's access to care (249/524, 47.5%), their desire not to practice or train in a state with abortion restrictions regardless of current residence (249/524, 47.5%), and their personal belief that abortion bans are human rights and/or body autonomy violation (197/524, 37.6%). Some respondents stated that the *Dobbs* decision would not impact their choice of practice location (41/524, 7.8%), and some supported it (35/594, 6.7%).

XSL•FO
RenderX

Conclusions: This study shows that abortion restrictions are having an impact on the practice location preferences of the physician workforce due to both patient care and personal factors. It is important that state policy makers and others who are considering abortion restrictions also consider how to address these concerns of physicians and medical students, to avoid worsening geographic maldistribution of physicians and worsening access to care from physicians for their citizens.

(Interact J Med Res 2025;14:e55035) doi:10.2196/55035

KEYWORDS

abortion; physician workforce; social media; reproductive health; medical education; abortion access; education; survey study; students; training; patient care; care; medical students; human rights; autonomy

Introduction

The US landmark *Dobbs v Jackson Women's Health Organization* decision allowed for widespread restrictions on abortion care, with 14 US states now enforcing total abortion bans and 27 more with bans based on gestational age [1,2]. These include Targeted Regulation of Abortion Providers (TRAP) laws that hamper and criminalize the practice of abortion [2].

While evidence affirms that abortion restrictions have deleterious effects on patient care and public health [3-6], it is important to understand that such policies also impact the health of physicians. A majority of physicians and medical students plan to build families during or after medical training, with thousands desiring pregnancy each year [7,8]. Many rely on infertility treatments, which abortion restrictions hamper [9]. Abortion restrictions, therefore, may deny a significant proportion of the physician workforce comprehensive family planning services, placing them at risk of forced birth [10]. Furthermore, they may also create moral injury among physicians from conflict between personal and professional morals, uncertainty regarding allowable practices, and fear of prosecution [11]. Those who provide abortion care may face increased stigma or even criminalization, depending on the state in which they train or practice. Those who are in restricted states and are not able to provide abortion care may struggle to navigate what is right for their patients versus what is legal, potentially worsening burnout and compassion fatigue [11,12].

Recent analysis from the American Association of Medical Colleges (AAMC) shows that fewer US MD seniors applied for residency positions in abortion-banned states versus nonban states in 2023 [13]. This includes a small but significant decline in the number of applications to obstetrics and gynecology residency programs in restrictive states in 2022 and 2023 [14]. To date, no study has described why physicians hold such preferences. Using an inductive analysis of free-response survey questions from our previous survey, this study aims to describe how state abortion restrictions may influence physicians' and students' decisions about where to live and practice.

Methods

Overview

We conducted a web-based, cross-sectional study for 2 weeks in August 2022. A nonprobabilistic sample of physicians (practicing physicians, fellows, and residents) and medical students were recruited from dedicated physician communities on social media (Twitter [rebranded as X in July 2023], Facebook, and Instagram [Meta Platforms]) through platforms like the American Medical Women's Association and Inside The Match. All physicians and medical students in the United States were eligible to participate, including both those who practice or intend to practice in reproductive health care and those who do not. There was no minimum age for participation. Physicians completed a questionnaire about the impact of overturning *Roe v Wade* on practice location preferences [15]. Respondents reported demographic information and their location preferences for residency (medical students) or fellowship and jobs (physicians). No identifying information was collected.

This analysis focused on the study respondents' stated practice location preferences. Quantitative data from this study were previously published [15]. Survey respondents were offered a free-response question, "Please share your thoughts about the overturning of *Roe v Wade* and how it will affect your decision about your (residency/job or fellowship) programs." An inductive thematic analysis was used [10]. We consulted the Standards for Reporting Qualitative Research to report the study findings [16]. The free-response item was included to allow respondents to contextualize their practice location preferences [17]. The study team is comprised of a medical student pursuing obstetrics and gynecology (OBGYN; SMM), residents in radiation oncology (MSL) and OBGYN (SF), a fellow in Complex Family Planning (AL), and practicing physicians in psychiatry (SAB and JAG) and internal medicine (SJ and VMA). Some team members practice or are training in locations with abortion restrictions, and some practice in less restrictive locations. A total of 4 independent coders (MSL, SAB, SMM, and SF) coded responses until thematic saturation was reached (n=73 for medical students and n=102 for residents, fellows, and practicing physicians) and established the code book through consensus discussion. After establishing the code book, 2 authors coded all responses (n=524), and differences were resolved by discussion. Statistics were done in IBM SPSS (version 29), and group comparisons were calculated with chi-square testing. The CHERRIES checklist for the reporting of internet surveys guided the reporting of the study (Multimedia Appendix 1) [18].

Ethical Considerations

The study was approved as exempt from review by the Institutional Review Board at the University of Chicago (IRB22-1066). Participants provided consent with the opportunity to opt out of the study and were not compensated

XSL•FO
RenderX

for participation. Data were collected without identifiers and are only accessible to study team members.

Results

Demographics

Of the 2063 survey respondents, 524 (25.4%) completed the free-response item. Respondents consisted of medical students (n=219), residents and fellows (n=129), and practicing physicians (n=176). Most identified as cisgender women (391/524, 74.6%). The majority (453/524, 86.5%) of respondents were of reproductive age (less than age 44) and had no children (361/524, 68.9%). Approximately half (261/524, 50.5%) resided in states where abortion bans were in place or anticipated; half (256/524, 49.5%) resided in states where abortion remains legal [19]. Roughly a fifth (114/524, 21.8%) specialized in OBGYN, 13.2% (69/524) specialized in family medicine, and 65.1% (341/524) specialized in another field. The complete demographics of the sample who answered the free-response portion appear in Table 1.

Respondents who answered the free-response item were similar to those who did not by gender (P=.07), race (P=.13), or whether they intended to provide abortion care (P=.22). Respondents in states with restrictive abortion bans (50.5%) were more likely to respond (P<.001) compared with those in a state without restrictive abortion bans (41.7%).

Free-response rates suggest that these qualitative data appropriately represent the spectrum of views on abortion rights and access. Of the overall sample, 82.3% (1698/2063) indicated they would prefer to apply where abortion access is preserved; among them, 23.1% (393) answered the free-response item versus 76.9% (1305/2063) who did not (P<.001). However, of the 9.7% (200/2063) who did not prefer to apply where abortion access was preserved, 41.5% (83) provided a free response, while 58.5% (117) did not (P<.001). Of the 11.1% (229) who indicated that abortion restrictions do not impact their preferences, 32.8% (75) responded versus 67.2% (154) who did not (P<.001).

XSL•FO
RenderX

Table 1. Demographics of medical students (n=219), residents and fellows (n=129), and practicing physicians (n=176) who answered the free response item.

Characteristic	Total (n=524), n (%)	Medical students (n=219), n (%)[a]	Residents and fellows (n=129), n (%)	Practicing physicians (n=176), n (%)
Gender[b]				
Woman	391 (74.6)	158 (72.1)	94 (72.9)	139 (79)
Man	109 (20.8)	48 (21.9)	30 (23.3)	31 (17.6)
Transgender and/or gender nonconforming	7 (1.4)	4 (1.9)	2 (1.6)	1 (0.6)
Prefer to describe	12 (2.3)	2 (0.9)	0 (0)	3 (1.7)
Prefer not to answer	43 (2.1)	7 (3.2)	3 (2.3)	2 (1.1)
Ethnicity[c]				
Hispanic	45 (8.6)	27 (12.3)	10 (7.8)	8 (4.5)
Not Hispanic	456 (87)	181 (82.6)	114 (88.4)	161 (91.5)
Prefer not to answer	23 (4.4)	11 (5)	5 (3.9)	7 (4)
Race[c]				
American Indian, Alaska Native, Native Hawaiian, or other Pacific Islander	2 (0.4)	2 (1)	0 (0)	0 (0)
Asian	49 (9.4)	18 (8.2)	9 (7)	22 (12.5)
Black, African American, or African	37 (7.1)	19 (8.7)	16 (12.4)	2 (1.1)
Multiracial[d]	27 (5.2)	13 (5.9)	6 (4.7)	8 (4.5)
White	361 (68.9)	140 (63.9)	91 (70.5)	130 (73.9)
Prefer to describe	15 (2.9)	8 (3.7)	1 (0.8)	6 (3.4)
Prefer not to answer	33 (6.3)	19 (8.7)	6 (4.7)	8 (4.5)
Sexual orientation				
Bisexual	51 (9.7)	30 (13.7)	8 (6.2)	13 (7.4)
Gay or lesbian	19 (3.6)	8 (3.7)	4 (3.1)	7 (4)
Heterosexual	404 (77.1)	157 (71.7)	106 (82.2)	141 (80.1)
Queer, pansexual, and/or questioning	21 (4)	7 (3.2)	6 (4.7)	8 (4.5)
Don’t know	3 (0.6)	3 (1.4)	0 (0)	0 (0)
Prefer to describe	6 (1.1)	4 (1.8)	0 (0)	2 (1.1)
Prefer not to answer	20 (3.8)	10 (4.6)	5 (3.9)	5 (2.8)
Age range[e] **(years)**				
≤44	453 (86.5)	218 (99.5)	127 (98.4)	68 (38.6)
≥45	71 (13.5)	1 (0.5)	2 (1.6)	108 (61.4)
Relationship status				
Single	128 (24.4)	72 (32.9)	29 (22.5)	27 (15.3)
Partnered	125 (23.9)	89 (40.6)	29 (22.5)	7 (4)
Married	251 (47.9)	53 (24.2)	68 (52.7)	130 (73.9)
Widowed	2 (0.4)	0 (0)	0 (0)	2 (1.1)
Divorced	5 (1)	1 (0.5)	0 (0)	4 (2.3)
Other	5 (1)	1 (0.5)	1 (0.8)	3 (1.7)
Prefer not to answer	8 (1.5)	3 (1.4)	2 (1.6)	3 (1.7)
Children				
Yes	163 (31.1)	25 (11.4)	22 (17.1)	116 (65.9)

XSL•FO
RenderX

Characteristic	Total (n=524), n (%)	Medical students (n=219), n (%)[a]	Residents and fellows (n=129), n (%)	Practicing physicians (n=176), n (%)
No	361 (68.9)	194 (88.6)	107 (82.9)	60 (34.1)
Respondent's current state of residence, by anticipated abortion restriction[f]				
Ban or likely ban[g]	261 (50.5)	125 (58.4)	66 (51.2)	70 (40.2)
Legal[h]	256 (49.5)	89 (41.6)	63 (48.8)	104 (59.8)
Specialties				
Obstetrics and gynecology	114 (21.8)	51 (23.3)	28 (21.7)	35 (19.9)
Family medicine	69 (13.2)	30 (13.7)	17 (13.2)	22 (12.5)
All others	341 (65.1)	138 (63.0)	84 (65.1)	119 (67.6)

[a]Includes US medical students (n=188) and international medical graduates applying to US residency programs (n=31).

[b]Nationally, medical students are 47.9% female and 52.9% male, residents and fellows are 46.8% female and 53% male, and practicing physicians are 35.9% female and 64.1% male [20].

[c]Nationally, medical students are 0.2% American Indian or Alaska Native, 54.6% White, 21.6% Asian, 6.2% Black or African American, 5.3% Hispanic, 8% multiple races, and 3.5% other. Nationally, residents and fellows are 0.11% American Indian or Alaska Native, 48.9% White, 26.6% Asian, 6% Black or African American, 9.2% Hispanic, 4% multiple races, and 3.1% other. Nationally, practicing physicians are 0.1% American Indian or Alaska Native, 63.9% White, 19.2% Asian, 3.6% Black or African American, 5.5% Hispanic, 2% multiple races, and 5.6% other [20,21].

[d]Respondents who selected more than one option are considered multiracial for the purpose of this study.

[e]Age 15-44 years is defined as reproductive age per the Centers for Disease Control and Prevention [22].

[f]Includes all 50 states, Puerto Rico, and the District of Columbia. Excludes the 7 respondents who indicated "other" on their location [19].

[g]Alabama, Arizona, Arkansas, Florida, Georgia, Idaho, Indiana, Iowa, Kentucky, Louisiana, Michigan, Mississippi, Missouri, Montana, North Carolina, North Dakota, Ohio, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, Utah, West Virginia, Wisconsin, and Wyoming [19].

[h]Alaska, California, Colorado, Connecticut, Delaware, District of Columbia, Hawaii, Illinois, Kansas, Maine, Maryland, Massachusetts, Minnesota, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Vermont, Virginia, and Washington [19].

Overview of the Inductive Analysis of Free-Response Survey Answers

There were 2 groups of themes and 2 stand-alone themes. One group described how practice location decisions impact patient care (Table 2), and the other captured workforce-related concerns (Table 3). The remaining themes included no impact and antiabortion sentiment.

XSL•FO
RenderX

Table 2. Patient factors influencing decisions about practice location emerging from the inductive analysis of the following: for students applying to residency, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your residency application and ranking decisions below," and for fellows and practicing physicians, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your decision about your job or fellowship programs" among respondents (n=524).

Theme	Students (n=219), n (%)	Example quote	Physicians (n=305), n (%)	Example quote
Patient access to abortion (or reproductive care)	84 (38)	I'm horrified when I imagine taking care of a teenager who is being forced to carry out a pregnancy. I'm terrified of the burden of caring for a NICU filled with babies who were born despite having anomalies that make their short lives painful. I can only hope I'm not assaulted or become pregnant without the option to terminate.	165 (54)	I want to be able to support my patients to make good decisions about pregnancy. I need to be able to refer people if they need termination of pregnancy. It goes against my ethics to have to deprive someone of that option. I care foremost about my patients. If one of my patients died because she couldn't get an abortion, I wouldn't be able to live with myself.
Did not want politics to interfere with medical care decisions	45 (20)	I never want to be in a situation where I face disciplinary and/or legal consequences for reporting a patient who is miscarrying (spontaneous or induced), and with the current climate, I genuinely fear that we may be moving toward the criminalization of abortion in many places. That risk is not worth it to me when I could train in so many other places.	78 (25)	A politician is unable to grasp the grey areas of obstetric care and the heartbreaking scenarios we encounter. It is bad enough that hospital administrators police our obstetric practice; we do not need another non-medical person telling us how to practice evidence-based medicine.
Challenges of providing any reproductive care to patients with an abortion ban	41 (19)	I was previously set on Ob-Gyn, but I am now looking seriously at other fields because of the politics surrounding women's health care. I don't want to have to worry about legal repercussions for providing the best care to my patients. This has strongly turned me away from Ob-Gyn as a medical specialty.	65 (21)	As an abortion provider, I know that as much as I care about serving a population with unmet needs, the inevitability of burnout working in a place where abortion is severely limited would be too much.
Challenges of providing patient care that is not reproductive in nature	15 (7)	It will significantly impact the ability of every physician to provide care to their patients, regardless of their specialty, as many medical conditions are exacerbated by pregnancy status	47 (15)	I'm a dermatologist, and this affects our practice, too! We prescribe Accutane every day, and if a patient does become pregnant while on this drug due to contraceptive failure, we recommend termination. We prescribe lots of other teratogenic drugs as well for many different cutaneous diseases, especially methotrexate. I don't know how I can practice in a state where pharmacists might refuse to fill MTX.

XSL•FO
RenderX

Table 3. Practice location decisions that are workforce-related emerging from the inductive analysis of the following: for students applying to residency, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your residency application and ranking decisions below," and for fellows and practicing physicians, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your decision about your job or fellowship programs" among respondents (n=524).

Theme	Students (n=219), n (%)	Example quote	Physicians (n=305), n (%)	Example quote
Not choosing to practice or train in a state with abortion restrictions	77 (35)	This decision has heavily affected my residency application process. Amazing programs that I've highly considered are now at the bottom of my list.	172 (56)	Just finished residency and specifically did not even consider jobs in states that ban the full spectrum of reproductive healthcare or states that looked like they would consider a ban. Overturning of Roe made me basically have to ignore half the country during my search. But given the job market today, finding a position in a state that allows me to actually care for my patients wasn't hard.
Personal belief that an abortion ban is a human rights/body autonomy violation	63 (29)	States that do not respect basic human rights are not places I wish to live or raise a family.	134 (44)	The overturning of Roe is the overturning of basic freedoms, the right to privacy, and bodily autonomy. It is the first step in overturning other rights. It is removing science from medicine. It threatens all doctors whether they provide abortion care or not. I'm likely to leave medicine, then practice in that environment and take those risks.
Access to training and education in abortion	43 (20)	I want to be part of a program where abortion training is easily accessible, and I will not have to go out of state to get this training. I also want to protect these rights for myself and my future patients.	15 (5)	One of my biggest decisions in choosing my state of residency was to allow me every opportunity to learn about women's care at all levels. The overturning will prevent students and residents from reaching their full potential of learning care for women. It is truly unfortunate that men outside of the walls of understanding of medical knowledge think they have the authority to control not only women's bodies but also the education of those to be able to treat women in emergency settings safely and holistically.
Personal or family access to abortion care or family building	36 (16)	I'm a guy, but what about my daughters in the future? What about a pregnancy complication with my wife? What about my patients? This is the problem when people claim moral high ground on the basis of their religion and are placed into positions of power; you end up with a sort of theocracy.	58 (19)	I was planning on looking for underserved community jobs in Idaho, but now that they have an early abortion ban, I will not be. I am actively trying to get pregnant and won't risk my life to pursue a job.
Geographic ties to states with abortion restrictions limiting relocation	18 (8)	I attend medical school in my home state, which hasn't banned abortion as of yet but might do so in the future. If abortion is banned here, I'll likely still rank in-state programs due to the proximity of my family, but I will not rank out-of-state programs where abortion is banned.	38 (12)	Unfortunately, my answers are influenced by the fact that I live in a state with some of the most restrictive policies and have no ability to move. I cannot simply uproot my life to another state due to my feelings on abortion access. I work here, and my husband works here. My family is here. His family is here. The best I can do is to advocate for change, but I must remain in place as the primary breadwinner in my family.
Challenges recruiting to states with abortion restrictions	0 (0)		18 (6)	I'm a program director and am concerned about how this will affect recruiting talented and eager physicians to our state. Our patients already have difficulty accessing the medical system, so if this decision leads to physicians leaving the state, it will only amplify disparities.

Patient Factors Influencing Decisions About Practice Location

Patient Access to Abortion or Full-Spectrum Reproductive Care

Many physicians and medical student respondents want patients to have access to safe and legal abortion. Respondents specifically highlighted concerns that adolescents, underrepresented minority groups, people in rural communities, and lower-income patients would increasingly face challenges in finding abortion providers, exacerbating health disparities (Table 2).

Physicians also noted that restrictions interfere with their ability to provide or refer patients for abortion care. For example, one stated, "I won't practice in a state that limits my ability to provide or refer my patients for care that is safe and necessary for their health and well-being."

XSL•FO
RenderX

Challenges of Providing Reproductive Care to Patients During an Abortion Ban

Reproductive health care providers anticipate moral distress if they are unable to provide abortion care in circumstances like lethal fetal anomalies or pregnancies resulting from rape or incest. An OBGYN physician explained, “Abortion care and prenatal care go hand in hand. This is a field with a lot of gray areas, and elimination of options will harm those who can get pregnant.” Many physicians feared legal repercussions and were disappointed by a perceived lack of institutional support for evidence-based health care.

Do Not Want Politics to Interfere With Medical Care Decisions

Some respondents expressed concern that lawmakers are interfering with medical care. Others emphasized the role physicians play in advocacy and supporting elected officials in favor of essential reproductive health care. A participant stated, “The government should have no standing in a medical decision between physician and patient.”

Challenges of Providing Patient Care That Is Not Reproductive in Nature

Physicians across various fields were concerned that abortion restrictions would adversely impact their clinical practice. For example, a pediatrician noted, “Working with fetal cardiac patients, it is imperative that my patients have access to abortion services if that’s the choice they make that’s best for their families.” In addition, an oncologist worried about restrictions on chemotherapy regimens, a dermatologist had questions about prescribing common medications (like Accutane) that are teratogenic, and a rheumatologist had concerns about prescribing methotrexate.

Workforce-Related Practice Location Decisions

Choosing Not to Practice or Train in a State With Abortion Restrictions

Many respondents living in states with abortion protections stated that they would be unwilling to move to a state with abortion restrictions (Table 3). Others living in restrictive states intend to move or preferentially apply to and rank training programs in states without abortion bans. Trainees described how these decisions compound their stress regarding the highly competitive match process. Some still felt pressured to apply everywhere, regardless of their personal preferences, stating, “Residency is already so competitive, so unfortunately, I feel like I have to apply everywhere, but I would definitely preferentially rank somewhere that I would have access to abortion care and that my patients would as well.”

Challenges Recruiting to States With Abortion Restrictions

Some residency and fellowship program directors and administrative leadership in states with restrictive abortion laws are concerned about recruiting and retaining residents, fellows, and faculty. Many foresee the reluctance of trainees and faculty to work in restrictive states. A program leader said, “I am an APD at an academic medical center in the Midwest. I have already been told by two residents that they had planned to stay in the state to practice but are now leaving solely because of the lack of reproductive rights in our state. I fear we will rapidly lose amazing physicians.”

Personal Belief That Abortion Restrictions Violate Human Rights and/or Bodily Autonomy

A substantial portion of respondents described the overturning of *Roe v Wade* as a human rights violation and criticized its negative impact on patients’ bodily autonomy. Others discussed the potential moral injury from practicing in a state whose laws and policies prevent clinicians from providing evidence-based medical care.

Respondents connected states’ abortion policies to their overarching sociopolitical climates, noting that bans and restrictions may portend other harmful (eg, racist, homophobic, transphobic) policies. A medical student said, “Extremely cautious about applying to these states who have denied abortion care. Not only because of abortion care but also because these states are notoriously anti-LGBTQ+ and hold racist values. I do not want to live and work and raise a family in that environment, where I am not respected and have less human rights than others.”

Access to Abortion Training and Education

Students applying to OBGYN and family medicine expressed that their application decisions would be shaped by access to proper training in abortion care. Applicants to residency and fellowship recognize that selecting programs in abortion-restricted states may limit access to adequate training. Multiple students noted that they intend to inquire about abortion training during the residency application process.

Some recognized that trainees in abortion-restricted states could seek abortion training out-of-state. For example, a respondent said, “I plan to first rank programs in states with full spectrum reproductive health access, followed by programs that are intentional about providing training for their residents with full support (financial, housing, etc) to leave the state for abortion training.” However, current trainees also discussed challenges in obtaining abortion training, including professional, administrative, and financial barriers.

Geographic Ties to States With Abortion Restrictions Limiting Relocation

Some noted that geographic relocation is a privilege not afforded to everyone equally. The decision to move is often influenced by distance to a support network, job benefits for the respondent or their spouse, housing, and childcare. Such geographic ties discourage or prevent many medical students and physicians from leaving their state of residence despite their personal or professional opposition to abortion restrictions.

Some said they understand the risks of staying in a state with abortion restrictions. If necessary, they would travel out of state to receive an abortion, again recognizing their mobility is a privilege. A respondent said, “I definitely would prefer to be in a state that maintains access to abortion. Unfortunately, those are not states where my family lives, and I am grateful that I

XSL•FO
RenderX

have enough privilege if I needed an abortion, I could leave the state."

Personal or Family Access to Abortion Care or Family Building

Respondents were concerned about practicing in a location that limits their options for receiving comprehensive reproductive health care. Multiple respondents highlighted that they did not want to be forced to carry a pregnancy if they could not get an abortion, especially during training. Others specifically cited medical conditions that would make pregnancy physically challenging and even contraindicated as a reason to ensure they had access to abortion care. A respondent said, "I am a medical student with chronic conditions that make pregnancy life-threatening for me. Although I am on contraceptives, nothing is 100%, and I want to be able to protect my life and well-being in case I do accidentally get pregnant."

In addition, physicians with infertility undergoing in vitro fertilization noted that practicing in a state where life is defined as beginning at fertilization would make family building significantly more challenging. Commonly, respondents stated they were concerned about care for themselves, their children, or their partners, underscoring the importance of recognizing that physicians, too, need access to care.

Additional Themes

No Impact

Few medical students and physicians stated that the *Dobbs* decision would not impact their choice of practice location (Table 4). Some indicated that the residency and fellowship match were too competitive to make decisions based on abortion legislation. For example, those who apply to every program in their field may end up applying to programs in states with abortion restrictions to increase their likelihood of matching.

Table 4. Practice location decisions that are workforce-related emerging from the inductive analysis of the following: for students applying to residency, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your residency application and ranking decisions below," and for fellows and practicing physicians, "Please share your thoughts about the overturning of Roe v Wade and how it will affect your decision about your job or fellowship programs" among respondents (n=524).

Theme	Students (n=219), n (%)	Example quote	Physicians (n=305), n (%)	Example quote
No impact	30 (14)	Matching and getting into a program is challenging enough considering the various factors at play; this decision will not be part of deciding which states or programs I end up applying to.	11 (4)	It will have zero impact on my decisions regarding jobs/fellowships.
Expressed support for overturning *Roe v Wade*	17 (8)	The overturning of Roe v. Wade is long overdue. It was not right in the first place, as the Supreme Court made clear in its ruling. Babies deserve to live inside and outside the womb.	18 (6)	I am supportive of the overturn and believe it will be better for our patients and medical care to ban an inhumane practice like abortion. Human lives in the womb deserve protection just like all of our other patients at any age and ability.

Expressed Antiabortion Sentiment and/or Support for Overturning Roe v Wade

Physicians and medical students who expressed antiabortion ("pro-life") views supported the Supreme Court decision (Table 4). Multiple respondents noted that they would purposefully seek out practice environments where abortion restrictions existed. Reasons for this include not supporting abortion care for any indication, stating that they do not view abortion as health care, a desire to "preserve life," and a desire to "protect the unborn." Multiple respondents discussed that abortion is an issue that should be legislated at the state level.

Discussion

Principal Findings

Our study shows that abortion restrictions will have a substantial impact on the physician workforce in patient care and practice location decisions. The 3 most common themes were patient access to care, not choosing to practice or train in a state with abortion restrictions, and personal belief that an abortion ban is a human rights/body autonomy violation. This study enhances emerging literature about the impacts of abortion restrictions on the physician workforce, including physicians and medical students at all levels of training across all 50 states within both reproductive and nonreproductive health fields.

Respondents shared concerns that abortion restrictions will negatively impact their ability to provide high-quality, comprehensive reproductive care. This was evident among trainees who provide abortion care, like OBGYN residents, who expressed concerns about new or worsening barriers to obtaining foundational skills like first-trimester uterine aspiration at their primary institution [23,24]. OBGYN trainees also cited multiple barriers to obtaining foundational abortion care skills at their primary institutions. Some programs have created away rotation opportunities for residents unable to obtain comprehensive abortion training at their own institutions [25]. However, there are many barriers to these programs, including obtaining state-based medical licenses, getting funding and organizational affiliations in place, and disruptions to families when living in another state [25].

Even within nonreproductive health care fields, respondents shared concerns about the downstream effects of abortion restrictions on clinical training and practice. In the 2 weeks following the *Dobbs* decision, only 38.5% of a list of 187 societies across a wide variety of specialties had made a statement about the decision [26]. Respondents from specialties

XSL•FO
RenderX

that do not provide abortion care noted concern for restricted use of potentially abortifacient or teratogenic medications and worsening health among patients whose physical or mental health will be adversely impacted by restrictions.

Physicians and medical students also worried that abortion restrictions would deleteriously affect their personal health and well-being. Restrictions hold significant health implications for reproductive-age women, a large and growing demographic of the physician workforce [27]. Recent studies have reported that abortion is common among physicians, affirming that physicians, too, need safe, legal access to abortion [28]. This study informs future medical education and occupational health research by elevating trainees' and employees' concerns. As highlighted by the medical student responses on geographic ties and competitiveness of the match process, it is critical to recognize the multifactorial decision-making involved in where to complete residency training. While before the *Dobbs* decision, telehealth may have been able to bridge the gaps in access to abortion care, this is less likely to be possible in the current landscape [29]. Medical schools and hospitals, especially those in restrictive states, must recognize this and prepare to navigate the adverse health, financial, and legal repercussions their employees may face. Otherwise, disparate abortion access may increase health disparities within the physician workforce and threaten its diversity and resiliency [30].

If medical students do not want to practice in states with abortion restrictions in place, it is less likely that they will establish their practice in those locations. In 2022, 55.2% of those completing training established their practice in the same state where they completed residency [31]. The lack of physicians who are willing to practice in states with abortion restrictions can further poor health outcomes in maternity care deserts [32,33]. Idaho is a notable example, where 41% of OBGYN physicians consider leaving and cite restrictive abortion laws as a motivation [34,35]. Idaho has the lowest rate of physicians per 100,000 people in the entire country [34,35].

Some physicians stated that abortion restrictions would not impact them or that they support them. Notably, a subset of "no impact" responders shared that the scarcity of available positions, particularly within highly competitive specialties and for historically marginalized applicants, outweighs their personal opposition to abortion restrictions. Others acknowledged the futility of setting preferences since the match is ultimately complex and multifactorial.

Limitations

This study may be limited by self-selection bias, given its recruitment of medical students and practicing physicians on social media. Of the respondents who did not prefer to apply where abortion access was preserved, a substantial number (41.5%) provided a free response, indicating that we had a spectrum of views on abortion rights. Furthermore, this sample is focused on physicians and does not represent other health care workforce members who are likely also impacted by abortion restrictions.

Conclusion

The findings of this study captured responses to abortion restrictions before the 2023 Match cycle and provided context to the recent AAMC data showing that residency applications disproportionately decreased in restrictive states [13]. Narrative responses bolster our original quantitative data, affirming that access to full-spectrum reproductive health care was highly valued personally and professionally by most physicians [15].

This study shows that abortion restrictions are having an impact on the practice location preferences of the physician workforce due to both patient care and personal factors. It is important that state policy makers and others who are considering abortion restrictions also consider how to address these concerns of physicians and medical students, to avoid worsening geographic maldistribution of physicians and worsening access to care from physicians for their citizens.

Acknowledgments

Generative artificial intelligence was not used to aid in writing this manuscript.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly to protect the anonymity of participants due to the sensitive subject matter of this manuscript but are available from the corresponding author on reasonable request.

Authors' Contributions

MSL and SAB contributed equally as cofirst authors. SAB and MSL had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis. SAB, MSL, JAG, and VMA contributed to the concept and design. All authors contributed to the acquisition, analysis, or interpretation of data. SAB, MSL, SMM, JAG, and VMA contributed to the drafting of the manuscript. All authors contributed to the critical revision of the manuscript for important intellectual content. MSL performed statistical analysis. JAG and VMA contributed to administrative, technical, or material support. JAG and VMA performed supervision.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
The Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[PDF File (Adobe PDF File), 316 KB - ijmr_v14i1e55035_app1.pdf]

References

1. Nash E, Guarnieri I. Six months post-Roe, 24 US States have banned abortion or are likely to do so: a roundup. Guttmacher Institute. 2023. URL: https://www.guttmacher.org/node/300254/printable/print [accessed 2023-03-13]
2. Abortion laws by state. Center for Reproductive Rights. 2022. URL: https://reproductiverights.org/maps/abortion-laws-by-state/ [accessed 2023-03-13]
3. Sabbath EL, McKetchnie SM, Arora KS, Buchbinder M. US obstetrician-gynecologists' perceived impacts of post-Dobbs v Jackson state abortion bans. JAMA Netw Open 2024;7(1):e2352109 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.52109] [Medline: 38231510]
4. Keegan G, Francis M, Chalmers K, Hoofnagle M, Noory M, Essig R, et al. Trauma of abortion restrictions and forced pregnancy: urgent implications for acute care surgeons. Trauma Surg Acute Care Open 2023;8(1):e001067 [FREE Full text] [doi: 10.1136/tsaco-2022-001067] [Medline: 36744294]
5. Stevenson AJ, Root L. Trends in maternal death post-Dobbs v Jackson Women's Health. JAMA Netw Open 2024;7(8):e2430035 [FREE Full text] [doi: 10.1001/jamanetworkopen.2024.30035] [Medline: 39190311]
6. Singh P, Gallo MF. National trends in infant mortality in the US after Dobbs. JAMA Pediatr 2024:e244276. [doi: 10.1001/jamapediatrics.2024.4276] [Medline: 39432283]
7. Blair JE, Mayer AP, Caubet SL, Norby SM, O'Connor MI, Hayes SN. Pregnancy and parental leave during graduate medical education. Acad Med 2016;91(7):972-978. [doi: 10.1097/ACM.0000000000001006] [Medline: 26606722]
8. Stentz NC, Griffith KA, Perkins E, Jones RD, Jagsi R. Fertility and childbearing among American female physicians. J Womens Health (Larchmt) 2016;25(10):1059-1065 [FREE Full text] [doi: 10.1089/jwh.2015.5638] [Medline: 27347614]
9. Feinberg EC, Kawwass JF, Cedars MI. Roe v Wade and the threat to fertility care. Obstet Gynecol 2022;140(4):557-559. [doi: 10.1097/AOG.0000000000004928] [Medline: 35861359]
10. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. International Journal of Qualitative Methods 2017;16(1):160940691773384. [doi: 10.1177/1609406917733847]
11. Desai A, Holliday R, Wallis M, Thornewill A, Bahraini NH, Borges LM. Policy changes as a context for moral injury development in the wake of Dobbs v Jackson Women's Health Organization. Obstet Gynecol 2023;141(1):15-21. [doi: 10.1097/AOG.0000000000005009] [Medline: 36701606]
12. Martin LA, Debbink M, Hassinger J, Youatt E, Harris LH. Abortion providers, stigma and professional quality of life. Contraception 2014;90(6):581-587. [doi: 10.1016/j.contraception.2014.07.011] [Medline: 25131444]
13. Training location preferences of U.S. medical school graduates post Dobbs v Jackson Women's Health Organization decision. American Association of Medical Colleges. URL: https://www.aamc.org/advocacy-policy/aamc-research-and-action-institute/training-location-preferences [accessed 2023-04-16]
14. Hammoud MM, Morgan HK, George K, Ollendorff AT, Dalrymple JL, Dunleavy D, et al. Trends in obstetrics and gynecology residency applications in the year after abortion access changes. JAMA Netw Open 2024;7(2):e2355017 [FREE Full text] [doi: 10.1001/jamanetworkopen.2023.55017] [Medline: 38324311]
15. Bernstein SA, Levy MS, McNeilly S, Fishbach S, Jain S, Gold JA, et al. Practice location preferences in response to state abortion restrictions among physicians and trainees on social media. J Gen Intern Med 2023;38(10):2419-2423. [doi: 10.1007/s11606-023-08096-5] [Medline: 36823418]
16. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
17. O'Cathain A, Thomas KJ. "Any other comments?" open questions on questionnaires - a bane or a bonus to research? BMC Med Res Methodol 2004;4:25 [FREE Full text] [doi: 10.1186/1471-2288-4-25] [Medline: 15533249]
18. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
19. McCann A, Walker AS, Sasani A, Johnston T. Tracking the states where abortion is now banned. New York Times. URL: https://www.nytimes.com/interactive/2024/us/abortion-laws-roe-v-wade.html [accessed 2024-11-28]
20. Diversity in medicine: facts and figures 2019. American Association of Medical Colleges. URL: https://www.aamc.org/data-reports/workforce/report/diversity-medicine-facts-and-figures-2019 [accessed 2023-04-09]
21. ACGME Data Resource Book. Accreditation Council for Graduate Medical Education. URL: https://www.acgme.org/about-us/publications-and-resources/graduate-medical-education-data-resource-book/ [accessed 2023-04-09]
22. Women's reproductive health. Centers for Disease Control and Prevention. 2022. URL: https://www.cdc.gov/reproductive-health/women-health/?CDC_AAref_Val=https://www.cdc.gov/reproductivehealth/womensrh/index.htm [accessed 2023-04-09]
23. Steinauer JE, Turk JK, Pomerantz T, Simonson K, Learman LA, Landy U. Abortion training in US obstetrics and gynecology residency programs. Am J Obstet Gynecol 2018;219(1):86.e1-86.e6. [doi: 10.1016/j.ajog.2018.04.011] [Medline: 29655963]

XSL•FO
RenderX

24. Vinekar K, Karlapudi A, Nathan L, Turk JK, Rible R, Steinauer J. Projected implications of overturning Roe v Wade on abortion training in U.S. obstetrics and gynecology residency programs. Obstet Gynecol 2022;140(2):146-149. [doi: 10.1097/AOG.0000000000004832] [Medline: 35852261]
25. !!ATTN #OBGYN Residents:!! OHSU is committed to educating the next generation of #abortion providers. Residents with limited or no access to abortion training are invited to apply a Complex Family Planning elective at OHSU for the 2023-2024 academic year. @eeyanmiller. 2023 Mar 15. URL: https://twitter.com/eeyanmiller/status/1636103934434758658 [accessed 2023-04-09]
26. Levy MS, Fishbach S, Salles A, Arora VM. When do medical societies take a stand on abortion? J Gen Intern Med 2023;38(6):1559-1562 [FREE Full text] [doi: 10.1007/s11606-022-07867-w] [Medline: 36323828]
27. Physician specialty data report. American Association of Medical Colleges. URL: https://www.aamc.org/data-reports/workforce/report/physician-specialty-data-report [accessed 2023-04-09]
28. Levy MS, Arora VM, Talib H, Jeelani R, Duke CMP, Salles A. Abortion among physicians. Obstet Gynecol 2022;139(5):910-912. [doi: 10.1097/AOG.0000000000004724] [Medline: 35576350]
29. Koenig LR, Becker A, Ko J, Upadhyay UD. The role of telehealth in promoting equitable abortion access in the United States: spatial analysis. JMIR Public Health Surveill 2023;9:e45671. [doi: 10.2196/45671] [Medline: 37934583]
30. Nobel K, Luke AA, Rice WS. Racial disparities in pregnancy options counseling and referral in the US South. Health Serv Res 2023;58(1):9-18 [FREE Full text] [doi: 10.1111/1475-6773.14049] [Medline: 36068681]
31. Table C6. physician retention in state of residency training, by state. American Association of Medical Colleges. URL: https://www.aamc.org/data-reports/students-residents/data/report-residents/2020/table-c6-physician-retention-state-residency-training-state [accessed 2023-04-09]
32. DiPietro Mager NA, Zollinger TW, Turman JE, Zhang J, Dixon BE. Routine healthcare utilization among reproductive-age women residing in a rural maternity care desert. J Community Health 2021;46(1):108-116. [doi: 10.1007/s10900-020-00852-6] [Medline: 32488525]
33. Wallace M, Dyer L, Felker-Kantor E, Benno J, Vilda D, Harville E, et al. Maternity care deserts and pregnancy-associated mortality in Louisiana. Womens Health Issues 2021;31(2):122-129 [FREE Full text] [doi: 10.1016/j.whi.2020.09.004] [Medline: 33069560]
34. Cunningham M. Survey shows Idaho's maternal health doctors are leaving the state, or soon will. Idaho Capital Sun. 2023 Apr 7. URL: https://idahocapitalsun.com/2023/04/07/survey-shows-idahos-maternal-health-doctors-are-leaving-the-state-or-soon-will/ [accessed 2023-04-09]
35. State physician workforce data report. American Association of Medical Colleges. URL: https://store.aamc.org/downloadable/download/sample/sample_id/506/ [accessed 2024-11-28]

Abbreviations

AAMC: American Association of Medical Colleges
OBGYN: obstetrics and gynecology
TRAP: Targeted Regulation of Abortion Providers

Edited by T de Azevedo Cardoso; submitted 05.12.23; peer-reviewed by A Arbabisarjou, A Stephenson-Famy; comments to author 03.07.24; revised version received 25.08.24; accepted 22.10.24; published 07.01.25.

Please cite as:
Levy MS, Bernstein SA, McNeilly SM, Liberty A, Fishbach S, Jain S, Gold JA, Arora VM
Student and Physician Views of How the Dobbs Decision Affects Training and Practice Location Preferences: Cross-Sectional Questionnaire Study
Interact J Med Res 2025;14:e55035
URL: https://www.i-jmr.org/2025/1/e55035
doi:10.2196/55035
PMID:

©Morgan S Levy, Simone A Bernstein, Sarah M McNeilly, Abigail Liberty, Shira Fishbach, Shikha Jain, Jessica A Gold, Vineet M Arora. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 07.01.2025. 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

Readmission After Ischemic Stroke in Ningxia, China, From 2017 to 2021: Retrospective Cohort Study

Hua Meng[1,2*], PhD; Xingtian Wang[3*], MA; Dongfeng Pan[4], BM; Xinya Su[5,6], MA; Wenwen Lu[7], MA; Zhuo Liu[5,6], MA; Yuhui Geng[5,6], MA; Xiaojuan Ma[5,6], MA; Ting Pan[5,6], MA; Peifeng Liang[8], PhD

[1]Hubei Provincial Clinical Research Center for Alzheimer's Disease, Tianyou Hospital, School of Medicine, Wuhan University of Science and Technology, Wuhan, China

[2]Brain Science and Advanced Technology Institute, Wuhan University of Science and Technology, Wuhan, China

[3]Medical Record Statistics Department, General hospital of Ningxia Medical University, Yinchuan, China

[4]Department of Emergency Medicine, People's Hospital of Ningxia Hui Autonomous Region, Ningxia Medical University, Yinchuan, China

[5]School of Public Health, Ningxia Medical University, Yinchuan, China

[6]Ningxia Key Laboratory of Environmental Factors and Chronic Disease Control, Yinchuan, China

[7]Futian Center for Chronic Disease Control, Shenzhen, China

[8]Department of Medical Affair, People's Hospital of Ningxia Hui Autonomous Region, Ningxia Medical University, 301 Zhengyuan North Street, Yinchhuan, China

[*]these authors contributed equally

Corresponding Author:
Peifeng Liang, PhD
Department of Medical Affair, People's Hospital of Ningxia Hui Autonomous Region, Ningxia Medical University, 301 Zhengyuan North Street, Yinchhuan, China

Abstract

Background: Stroke remains a major cause of death and disability worldwide. Ischemic stroke is the most common type of stroke. Readmissions after hospitalization increase the patient burden and waste health resources.

Objective: This study aimed to calculate rehospitalization rates and explore risk factors associated with rehospitalization in ischemic stroke.

Methods: In this retrospective cohort study, we identified 12,782 patients admitted for ischemic stroke at People’s Hospital of Ningxia Hui Autonomous Region between January 2017 and December 2021. Groups were determined based on the ID number. The most important factors were selected using the Least Absolute Shrinkage and Selection Operator regression model. Stabilized inverse probability of treatment weighting (SIPTW) was used to correct baseline imbalances between groups. The adjusted hazard ratios and Kaplan-Meier survival curves of significant factors after SIPTW were calculated using stepwise backward Cox regression.

Results: A total of 10,727 patients were included in the study. Among them, 12.7% and 7.2% were readmitted within 5 years and 1 year, respectively. Stepwise backward Cox analysis of SIPTW showed that diabetes was the influencing factor for rehospitalization within 5 years (1.15, 1.02 - 1.30) and 1 year (1.21, 1.03 - 1.43). Additionally, the female gender was identified as a protective factor against readmission within 5 years (0.83, 0.74 - 0.93).

Conclusions: Although the rate of rehospitalization varied among patients with ischemic stroke at different time points, the significant factors remained consistent. Therefore, early prevention and treatment methods may be consistent.

(Interact J Med Res 2025;14:e67522) doi:10.2196/67522

KEYWORDS

ischemic stroke; rehospitalization; Least Absolute Shrinkage and Selection Operator; stabilized inverse probability of treatment weighting; cox proportional hazards regression model

Introduction

Because of its high mortality and disability rate, cerebrovascular disease (CVD) seriously threatens human health and brings great pressure to the medical care system, especially in limited-income countries [1,2]. CVD is the primary cause of death and disability in adults in China [3,4]. It is estimated that about 330 million patients experience difficulties from CVD in China [5]. The Annual Report on Cardiovascular Health and Diseases in China (2021) shows that, in 2019, the total number of discharged patients with CVD was 26.8441 million. The total hospitalization expenses of CVD were RMB 136.028 billion.

XSL•FO
RenderX

After adjusting for price factors, the average annual growth rate of total hospitalization expenses for ischemic stroke and hemorrhagic stroke has been 18.82% and 13.51%, respectively, since 2004 [6]. Readmissions are common: up to 22% of individuals experience 30-day readmissions after neurologic hospitalization [7].

High readmission rates may indicate unresolved problems at initial discharge [8], the quality of immediate post-hospital care, a more chronically ill population, or combinations of these factors. High readmission rates are also associated with a substantial economic burden on the health care system and may represent opportunities to reduce avoidable costs [9]. Reduction of readmission rates has become the goal of national health care reform, health insurance, and Medicaid service centers.

To reduce rehospitalization rates in patients with ischemic stroke, it is essential to fully understand preventable and unpreventable predictors that may influence rehospitalization. Previous studies have reported that infection [10], advanced age [11], and diabetes [12] are the most common causes of rehospitalization in patients with stroke. Furthermore, because most readmissions were measured within 30 days of the event, it is unknown whether the reasons for long-term readmissions differ [13].

This study focused on 3 main areas to better understand the factors contributing to rehospitalization in patients with ischemic stroke: the rates of rehospitalization at various intervals (1 year and 5 years), the differences in patient characteristics between those who were rehospitalized and those who were not, and the factors influencing rehospitalization in patients.

Methods

Ethical Considerations

This study followed the principles of the Declaration of Helsinki and was approved by People's Hospital of Ningxia Hui Autonomous Region (2020-KY-053), and the informed consent was waived off by the review boards due to the nature of this research. All data in the manuscript and supplementary materials were anonymized in accordance with ethical standards, ensuring no personally identifiable information could be discerned.

Data Resources and Patients

This is a retrospective cohort study. In 2022, we examined electronic health record data from all patients who were discharged from People's Hospital of Ningxia Hui Autonomous Region between January 2017 and December 2021. The electronic health record data were anonymized and accessed in a secure environment. The data have a hierarchical structure, including medical record number, demographic characteristics, primary and secondary diagnoses, procedures, method of payment, and a total of 642 variables. The hospital enforces rigorous follow-up and intervention protocols for patients with stroke. Qualified specialists in brain and heart health provide guidance on exercise and rehabilitation after discharge via the WeChat app. Public health physicians conduct telephone follow-ups at 3 and 6 months post-discharge, while a professional doctor performs an in-person visit at 12 months. This visit primarily includes blood sample testing, carotid ultrasound examinations, and adjustments to the patient's medication regimen.

The criteria for data included in this study were as follows: (1) patients aged 18 years or older with a principal diagnosis of ischemic stroke (International Classification of Diseases, Tenth Edition [ICD]-10: I63); (2) patients readmitted to any hospital after discharge due to an ischemic stroke-related condition. The first onset of ischemic stroke was considered the index event, and readmission was the endpoint event. Patients discharged from the hospital who were deceased, under 18 years of age, or had incorrect ID numbers and encoding formats were excluded. If the same patient had multiple rehospitalization records, only the first 2 hospitalization records were retained.

Outcome Measures

Rehospitalization for ischemic stroke refers to the same individual being readmitted to the hospital with ischemic stroke as the primary diagnosis. The rehospitalization rate was calculated by dividing the number of patients readmitted after hospitalization by the total number of patients discharged alive during the same period [14].

Rehospitalizations were identified using ID numbers and the main diagnostic code for the disease from the electronic health data of The People's Hospital of Ningxia Hui Autonomous Region from 2017 to 2021. The primary outcome was a 5-year rehospitalization rate, and the secondary outcomes were 1-year rehospitalization rates.

Study Variables

Published relevant literature was reviewed to summarize and integrate the factors involved [7,9,11,12,15-20]. Covariates considered for confounding adjustment included sex, age (<60, 60 - 69, 70 - 79, ≥80), admission route, length of stay (LOS), anemia (D50-64), thyroid disease (E00-07), diabetes (E10-14), dementia (F00-03), parkinsonism (G20), transient ischemic attack and related syndrome (G45), hypertension (I10-15), coronary heart disease (CHD) (I20-25), paroxysmal tachycardia (I47), atrial fibrillation and flutter (I48), heart failure (I50), arteriosclerosis (I70), embolism and thrombosis (I74), acute upper respiratory tract infection (J00-06), pneumonia (J12-18), renal failure (N17-N19), and urinary tract infection (N39.000), treatment (anticoagulants, thrombectomy, thrombolysis, thrombectomy and thrombolysis), and NIH Stroke Scale (NIHSS) score.

Statistical Analysis

The variables of admission route and NIHSS score had missing values; the number of missing values was 5 (0.04%) and 10346 (96.4%), respectively. The missing mechanism was determined to be missing completely at random according to the correlation coefficient matrix between the missing values and other variables (Table S1 in Multimedia Appendix 1) [21]. Therefore, the incomplete data for admission route and NIHSS score were imputed simultaneously using multiple imputations (n=25) with the R package MICE (TNO and University of Twente) [22,23]. Based on the Akaike information criterion (AIC) value, one of the imputed datasets was selected for analysis.

XSL•FO
RenderX

Some previous studies have demonstrated that the Least Absolute Shrinkage and Selection Operator (LASSO) method was superior to traditional methods [24-26]. LASSO regression was used to avoid overfitting and collinearity [26]. Therefore, LASSO analysis was used to select variables to be included in the Cox proportional hazards regression model (Cox model). The "glmnet" package was used to analyze the LASSO regression model [27].

Stabilized inverse probability of treatment weighting (SIPTW) was used to achieve a balanced comparison between the readmission and non-readmission groups. This method helps maintain the original data's sample size and ensures an appropriate class I error rate [28]. Probability was estimated through a logistic regression model with rehospitalization as the dependent variable, considering variables such as age, anemia, hypertension, CHD, length of hospital stay, treatment, urinary tract infection, and NIHSS [29]. The balance of potential confounders at baseline was evaluated using the absolute standardized difference (SMD), where an SMD greater than 0.1 indicated a significant difference in potential confounders between cases and controls [30].

The Cox model was constructed by backward Cox regression using the AIC selection criteria, and the best model was chosen based on the least AIC [31-33]. The "survival" package was used for Cox analysis, and the "MASS" package was used to perform stepwise backward analysis. The log-rank test conducted stepwise backward Cox analysis of significant factors for rehospitalization rates for SIPTW. The Kaplan-Meier curve was obtained using the "survminer" package.

Model Evaluation

Calibration curves were plotted using the "rms" package, which compares the agreement between the model's predicted and observed probabilities [34].

Decision curve analysis was conducted using the 'ggDCA' package to assess the utility of a model in supporting clinical decisions [35].

Clinical impact curves were drawn using the 'rmda' package to evaluate the model's recognition value in rehospitalized patients [36].

A 2-sided *P* value of less than .05 indicated statistical significance. Data screening and extraction were performed using Excel version 2016, and other analyses were carried out using R (version 4.2.2).

Results

The results of the data inclusion procedure are shown in Figure 1. A total of 10,727 eligible patients with ischemic stroke were included in the study. Among all patients, 12.7% (1367) were readmitted within 5 years after discharge, and 7.2% (769) were rehospitalized within 1 year. The Kaplan-Meier curves are shown in Figure 2.

There was a significant imbalance in age, treatment, anemia, hypertension, CHD, and LOS between the 2 groups with or without rehospitalization within 5 years. After using the SIPTW, the SMD did not exceed 0.1. Baseline characteristics are shown in Table 1. Significant differences in age, anemia, hypertension, CHD, diabetes, LOS, and NIHSS were observed between the 2 groups with or without rehospitalization within 1 year. Following SIPTW adjustment, balance was achieved between the 2 groups (Table S2 in Multimedia Appendix 1).

On readmission within 5 years, sex, admission route, diabetes, heart failure, acute upper respiratory tract infection, paroxysmal tachycardia, LOS, embolism, and thrombosis were indicated for inclusion from a 5-year SIPTW LASSO regression. Analyzing the above variables mentioned above, stepwise backward Cox regression after SIPTW showed that the significant variable was diabetes, with a hazard ratio (HR) of 1.15 and a 95% confidence interval (95% CI) of 1.02 to 1.30. As shown in Figure 3, among 3175 patients with ischemic stroke with diabetes, 436 (13.7%) patients with ischemic stroke and diabetes were rehospitalized for treatment within five years. Sex (HR 0.83; 95% CI 0.74 - 0.93) was identified as a protective factor for rehospitalization in patients with ischemic stroke (Table 2). The Kaplan-Meier curves are displayed in Figure 3(B and C).

Sex, diabetes, paroxysmal tachycardia, and urinary tract infection were indicated for inclusion from a 1-year SIPTW LASSO regression (Figure S1 in Multimedia Appendix 1). In the stepwise backward Cox analysis of SIPTW, the significant factor was diabetes (HR 1.21; 95% CI 1.03 - 1.43).

Unweighted stepwise backward Cox analyses are listed in Table S3 in Multimedia Appendix 1.

XSL•FO
RenderX

Figure 1. Flowchart of patient selection for this study.

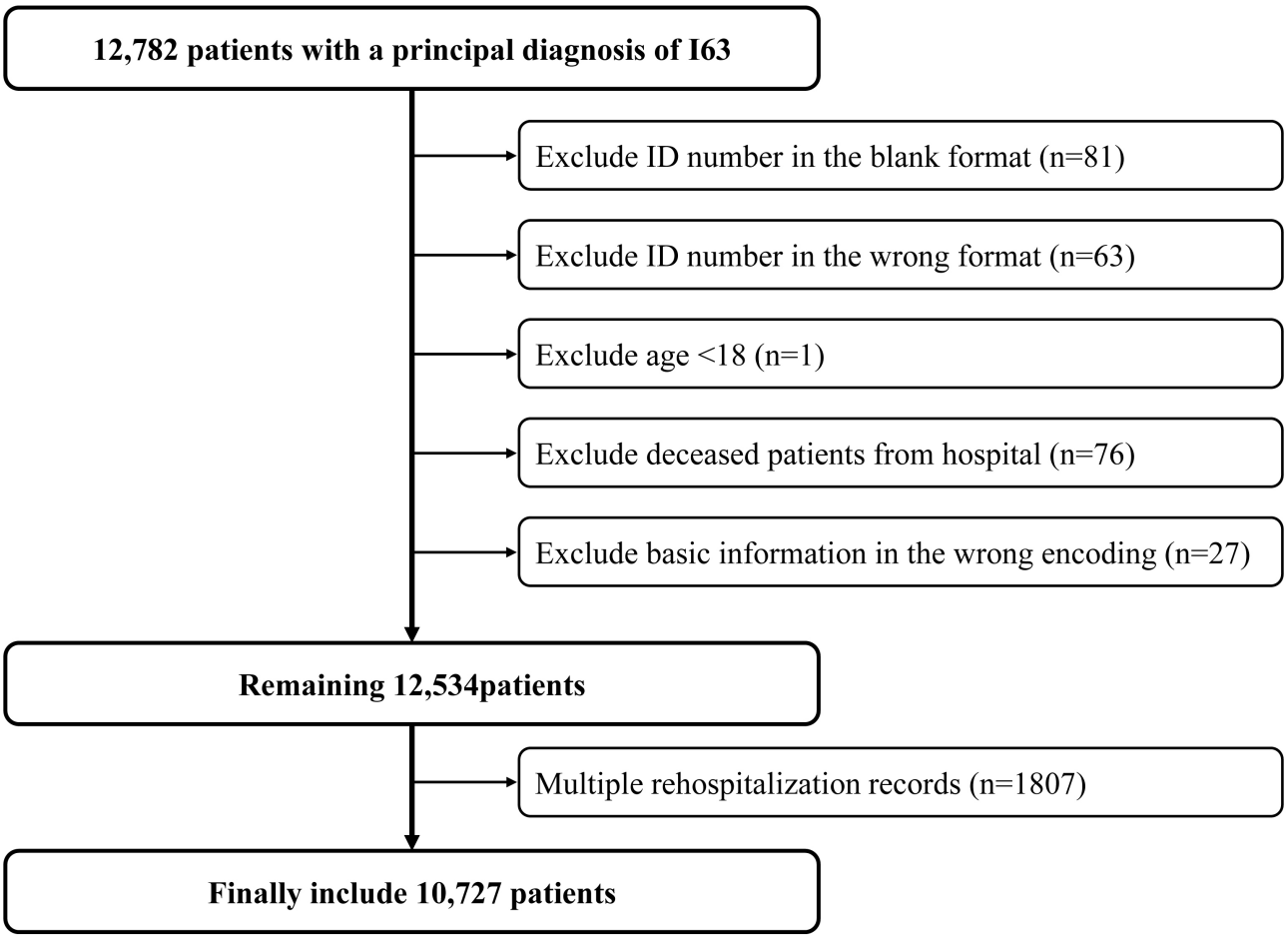
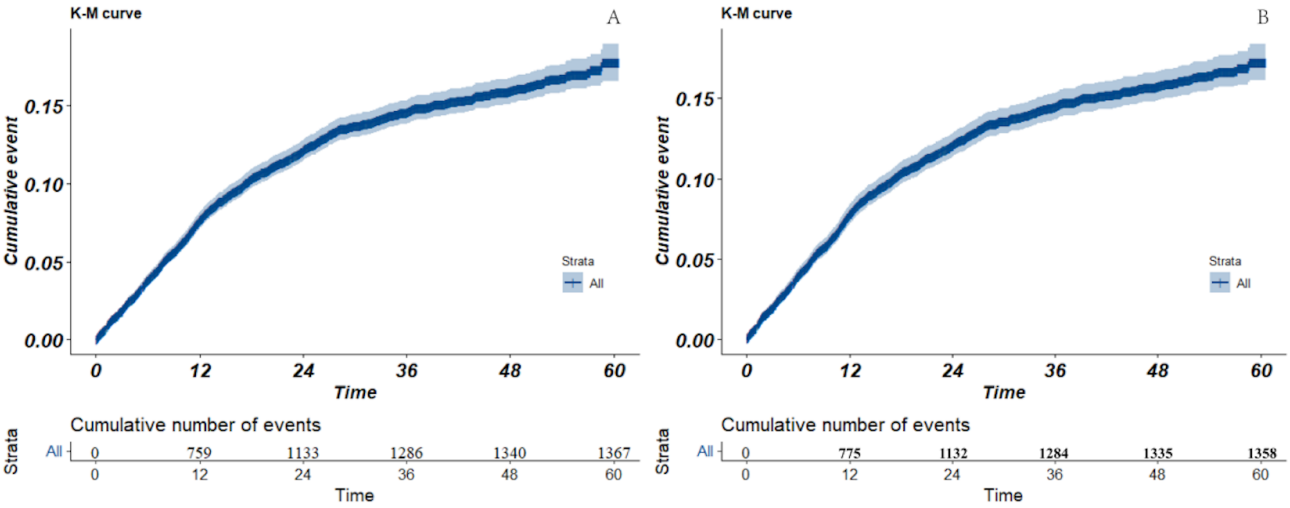


Figure 2. Kaplan-Meier curve of rehospitalization patients. A: cohort data; B: stabilized inverse probability of treatment weighting data. K-M curve: Kaplan-Meier curve.



XSL•FO
RenderX

Table . Characteristics between readmission and no readmission groups within 5 years.

	Cohort data				SIPTW[a] data			
	No readmission	Readmission	*P* value	SMD	No readmission	Readmission	*P* value	SMD
	N=9360	N=1367			N=9360.8	N=1358.1		
Male (n, %)	5159 (55.1)	794 (58.1)	.039	0.060	5148.3 (55.0)	797.0 (58.7)	.020	0.075
Age, years (n, %)								
<60	2267 (24.2)	227 (16.6)	<.001	0.211	2175.2 (23.2)	305.5 (22.5)	.863	0.021
60 - 69	2650 (28.3)	375 (27.4)			2639.6 (28.2)	388.8 (28.6)		
70 - 79	2911 (31.1)	481 (35.2)			2960.1 (31.6)	427.1 (31.5)		
≥80	1532 (16.4)	284 (20.8)			1585.8 (16.9)	236.6 (17.4)		
Admission route (n, %)								
Emergency	4579 (48.9)	629 (46.0)	.187	0.060	4567.3 (48.8)	625.7 (46.1)	.259	0.065
Others	1 (0.0)	0 (0.0)			1.0 (0.0)	0.0 (0.0)		
Outpatient	4768 (51.0)	736 (53.8)			4780.2 (51.1)	728.4 (53.6)		
Transferred	12 (0.1)	2 (0.1)			12.2 (0.1)	3.9 (0.3)		
Treatment (n, %)								
Anticoagulants	8847 (94.5)	1328 (97.1)	<.001	0.133	8879.2 (94.9)	1295.3 (95.4)	.931	0.028
Thrombectomy	19 (0.2)	1 (0.1)			17.5 (0.2)	2.7 (0.2)		
Thrombectomy and thrombolysis	38 (0.4)	2 (0.1)			34.9 (0.4)	3.7 (0.3)		
Thrombolysis	456 (4.9)	36 (2.6)			429.2 (4.6)	56.4 (4.2)		
Disease (n, %)								
Parkinsonism (n, %)	111 (1.2)	25 (1.8)	.047	0.053	112.8 (1.2)	18.5 (1.4)	.586	0.014
Anemia (n, %)	367 (3.9)	28 (2.0)	.001	0.110	344.9 (3.7)	52.9 (3.9)	.786	0.011
Thyroid disease (n, %)	1249 (13.3)	190 (13.9)	.574	0.016	1253.0 (13.4)	190.6 (14.0)	.550	0.019
Dementia (n, %)	184 (2.0)	33 (2.4)	.272	0.031	186.5(2.0)	33.5 (2.5)	.300	0.032
Transient ischemic attack and related syndrome (n, %)	643 (6.9)	108 (7.9)	.163	0.039	647.1 (6.9)	111.2 (8.2)	.129	0.048
Hypertension (n, %)	6606 (70.6)	1045 (76.4)	<.001	0.133	6677.2 (71.3)	973.1 (71.7)	.835	0.007
Coronary heart disease (n, %)	2132 (22.8)	383 (28.0)	<.001	0.121	2196.0 (23.5)	327.4 (24.1)	.616	0.015
Paroxysmal tachycardia (n, %)	660 (7.1)	68 (5.0)	.004	0.087	668.3 (7.1)	65.3 (4.8)	.003	0.099

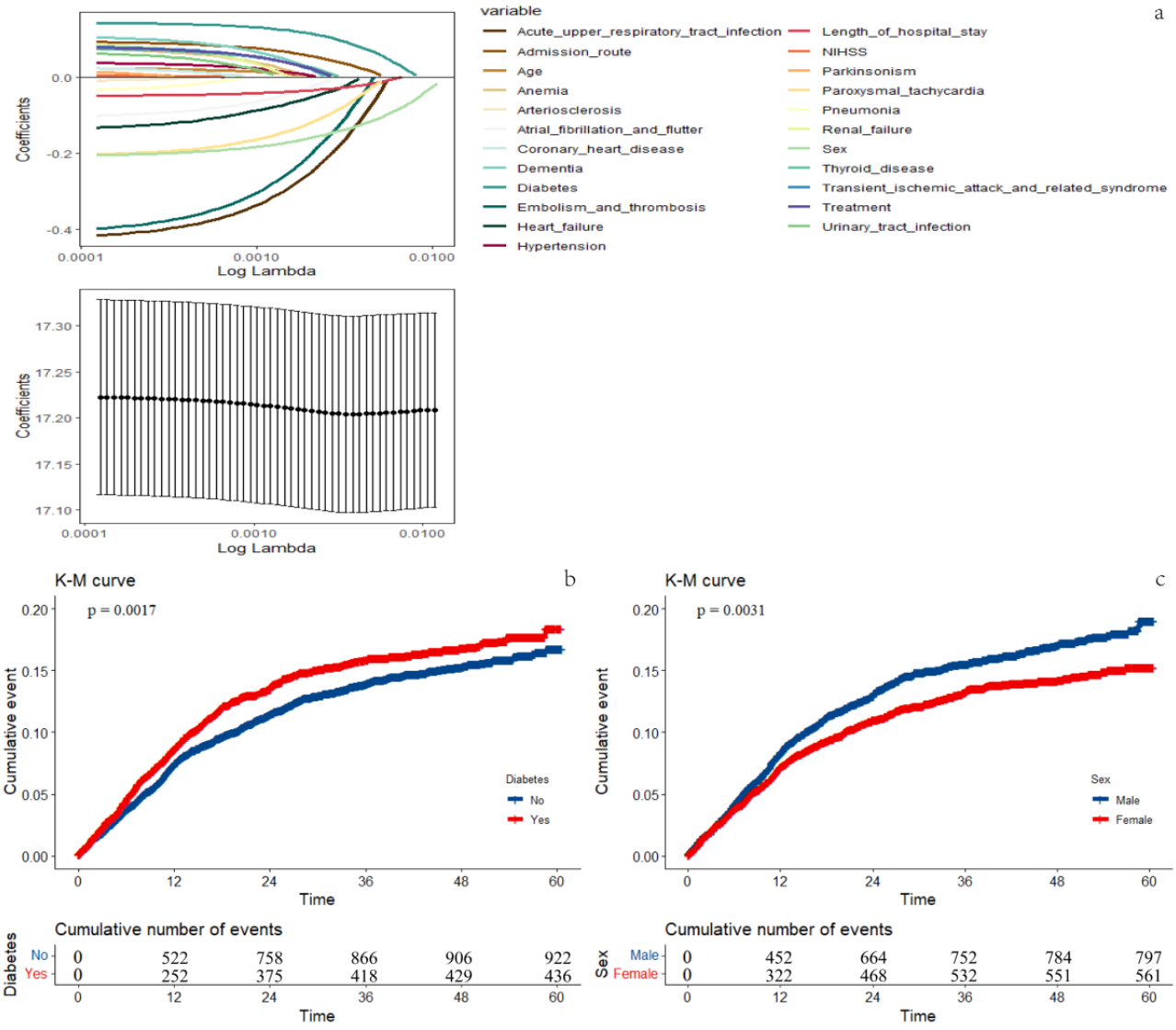
XSL•FO
RenderX

	Cohort data				SIPTW[a] data			
	No readmission	Readmission	*P* value	SMD	No readmission	Readmission	*P* value	SMD
	N=9360	N=1367			N=9360.8	N=1358.1		
Diabetes (n, %)	2706 (28.9)	446 (32.6)	.005	0.081	2738.5 (29.3)	435.6 (32.1)	.052	0.061
Atrial fibrillation and flutter (n, %)	564 (6.0)	72 (5.3)	.267	0.033	570.5 (6.1)	67.4 (5.0)	.124	0.050
Heart failure (n, %)	735 (7.9)	122 (8.9)	.172	0.039	758.5 (8.1)	102.2 (7.5)	.476	0.022
Arteriosclerosis (n, %)	5190 (55.4)	812 (59.4)	.006	0.080	5221.6 (55.8)	792.1 (58.3)	.113	0.051
Embolism and thrombosis (n, %)	165 (1.8)	18 (1.3)	.234	0.036	167.3 (1.8)	14.8 (1.1)	.074	0.059
Acute upper respiratory tract infection (n, %)	127 (1.4)	11 (0.8)	.091	0.053	128.5 (1.4)	11.9 (0.9)	.192	0.047
Pneumonia (n, %)	309 (3.3)	43 (3.1)	.763	0.009	315.1 (3.4)	47.2 (3.5)	.869	0.006
Renal failure (n, %)	148 (1.6)	20 (1.5)	.742	0.010	148.7 (1.6)	21.0 (1.5)	.915	0.004
Urinary tract infection (n, %)	324 (3.5)	32 (2.3)	.031	0.067	310.4 (3.3)	42.8 (3.2)	.815	0.009
Length of hospital stay (n, %)								
<Q1 (<8)	2394 (25.6)	162 (11.9)	<.001	0.414	2229.9 (23.8)	311.2 (22.9)	.545	0.023
Q1-Q2 (8 - 10)	2029 (21.7)	262 (19.2)			1999.3 (21.4)	291.2 (21.4)		
Q2-Q3 (10 - 13)	2728 (29.1)	440 (32.2)			2764.9 (29.5)	410.7 (30.2)		
≥Q3 (≥13)	2209 (23.6)	503 (36.8)			2366.7 (25.3)	344.9 (25.4)		
NIHSS (n, %)								
Q1 (≤1)	2459 (26.3)	389 (28.5)	.008	0.100	2483.9 (26.5)	351.6 (25.9)	.476	0.023
Q1-Q2 (1 - 2)	2766 (29.6)	361 (26.4)			2728.6 (29.1)	396.8 (29.2)		
Q2-Q3 (2 - 4)	2125 (22.7)	348 (25.5)			2159.7 (23.1)	325.1 (23.9)		
>Q3 (>4)	2010 (21.5)	269 (19.7)			1988.5 (21.2)	284.6 (21.0)		

[a]SIPTW: stabilized inverse probability of treatment weighting.

XSL•FO
RenderX

Figure 3. Stabilized inverse probability of treatment weighting and Least Absolute Shrinkage and Selection Operator (LASSO) regression, Kaplan-Meier (K-M) curves of rehospitalization within 5 years in patients with ischemic stroke (a): LASSO regression; (b) within 5 years in patients with ischemic stroke without or with diabetes; and (c) within 5 years in different sex patients with ischemic stroke. NIHSS: NIH Stroke Scale.



XSL•FO
RenderX

Table . Stepwise backward COX regression after SIPTW[a].

	B	SE	*P* Value	HR[b]	Lower limit	Upper limit
5 years						
Sex (female)	−0.188	0.055	.002	0.83	0.74	0.93
Diabetes (yes)	0.140	0.058	.027	1.15	1.02	1.30
Embolism and thrombosis (yes)	−0.388	0.262	.155	0.68	0.40	1.16
Paroxysmal tachycardia (yes)	−0.209	0.127	.130	0.81	0.62	1.06
Acute upper respiratory tract infection (yes)	−0.445	0.291	.185	0.64	0.33	1.24
1 year						
Sex (female)	−0.135	0.073	.096	0.87	0.74	1.02
Diabetes (yes)	0.192	0.077	.023	1.21	1.03	1.43
Paroxysmal tachycardia (yes)	−0.319	0.168	.107	0.73	0.49	1.07

[a]SIPTW: stabilized inverse probability of treatment weighting.

[b]HR: hazard ratio.

Model Validation

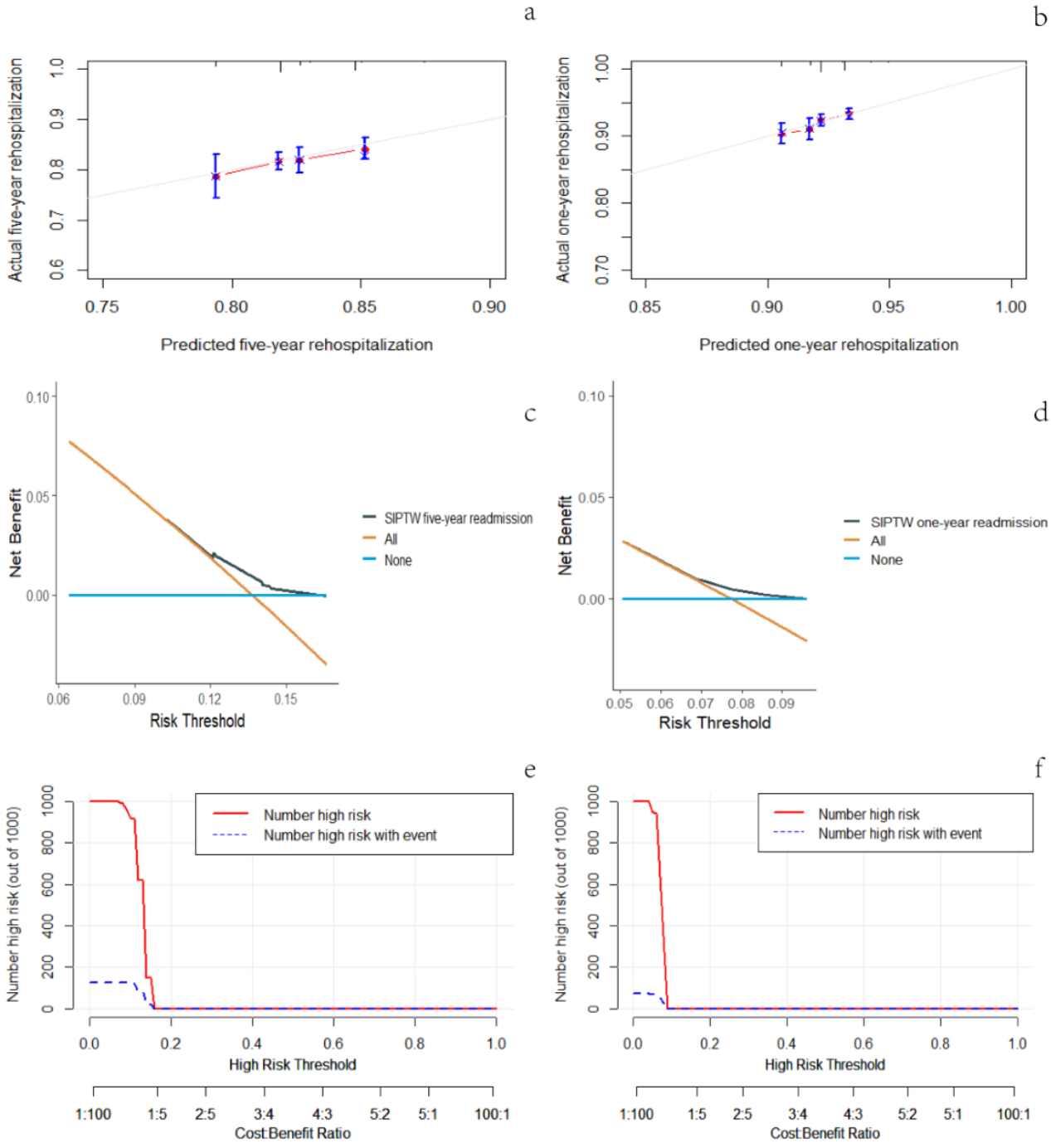
After performing SIPTW and stepwise backward Cox analysis, it can be observed from Figure 4(A and B) that the calibration curve of the model closely aligns with the diagonal line, suggesting that the model has good predictive power.

The decision curve analysis of the model indicated that if the patient's risk threshold probability for rehospitalization within 5 years was between 0.074 and 0.165 (Figure 4C), and within 1 year was between 0.057 and 0.096 (Figure 4D), then using the model to determine the need for rehospitalization offers added advantages over the options of full rehospitalization or no rehospitalization.

The clinical impact curve of the model showed that when the risk threshold was greater than 0.13, the high-risk rehospitalization cases of ischemic stroke predicted by the model within 5 years closely matched the true ischemic stroke rehospitalized cases (Figure 4E). When the threshold was above 0.07, the high-risk rehospitalization cases of ischemic stroke predicted by the model within 1 year closely resembled the true ischemic stroke rehospitalized cases (Figure 4F).

XSL•FO
RenderX

Figure 4. Model evaluation of results from 5-year and 1-year rehospitalization analysis after SIPTW (a, b: calibration curve; c, d: decision curve analysis; e, f: clinical impact curve. SIPTW: stabilized inverse probability of treatment weighting.



Discussion

Principal Findings and Comparison With Previous Works

In this study, there were similarities and differences in the influencing factors of rehospitalization at different times. The SIPTW and unweighted stepwise Cox analyses found gender to be significant among the LASSO regression factors for rehospitalization within 5 years.

A study in Sichuan Province included 1,066,752 patients with stroke with an average follow-up of 1.15 years, with a rehospitalization rate of 23% [37]. Between 1986 and 2001, there were 128,511 stroke hospitalizations in Scotland, and approximately 10.8% of patients were rehospitalized within 5 years due to stroke [38]. A study in Singapore included 12,559 patients with stroke, and the rehospitalization rate for recurrent stroke within 5 years was approximately 18.4% [39]. In our study, the low rate of rehospitalization may be attributed to the effective communication between medical staff and patients and their families upon discharge [40], which improved patient compliance with treatment and the professionalism of the accompanying staff. On the other hand, participants were from a single hospital, and in cases of acute illness, they may opt for medical care from nearby facilities like community general

XSL•FO
RenderX

practitioners or other urban areas in Ningxia or other provinces [41]. Additionally, these patients are typically managed by specialized palliative care teams and may choose not to be readmitted in the event of complications, opting instead for care at the palliative care facility. The COVID-19 pandemic also had an indirect impact on readmission rates, which is primarily reflected in the reduced accessibility of medical resources and changes in patient health care-seeking behavior. Our study did not systematically collect data on telemedicine usage during the pandemic, which may result in a potential underestimation of readmission rates. Future research should incorporate indicators of health care resource utilization to more comprehensively evaluate the long-term effects of the pandemic on the management of ischemic stroke.

Relevant meta-analyses have shown that common patient-related risk factors associated with increased readmission rates include age, heart failure, nephropathy, respiratory disease, peripheral arterial disease, and diabetes [42]. This aligns with the variables initially screened by LASSO regression in this study. In this study, the results of weighted stepwise Cox regression analysis and unweighted stepwise Cox regression analysis are inconsistent due to variations in the distribution of these variables between the non-rehospitalization group and the rehospitalization group in the original cohort data. Through the SIPTW method used in this study, the original dataset is preserved while adjusting for individual differences between groups, leading to a more balanced data distribution and more reliable results [43].

After adjusting for confounding variables, the study found that women patients had a lower risk of re-hospitalization within 5 years. In the analysis of both 5-year rehospitalization and 1-year rehospitalization, diabetes was considered a risk factor, which aligns with findings from previous studies [19,39,44]. This indicates that readmissions are more prevalent among men and individuals with diabetes, leading to higher health care costs [45]. Patients with diabetes exhibit a more pronounced procoagulant state [46] and experience delayed reperfusion of the ischemic penumbra [47], resulting in poorer recovery for patients with stroke. Experimental evidence suggests that hyperglycemia reduces the number of protective non-inflammatory macrophages, thereby increasing mediators of ischemic brain injury [48] and disrupting the blood-brain barrier [49], impacting the prognosis of patients with ischemic stroke.

Strengths

The main strength of this study is that we investigated patients with ischemic stroke at the People's Hospital of Ningxia Hui Autonomous Region over a 5-year period. We conducted follow-ups using electronic medical records from multiple hospitals. We included comprehensive details from the initial page of each patient's medical record for a comparative summary. Additionally, we used LASSO regression to screen factors, used SIPTW to adjust for confounding variables between groups, and conducted backward stepwise Cox analysis to refine the model.

Limitations

This study also has some potential limitations. First, we identified study subjects using ICD-10 diagnosis codes in the database. However, the coding accuracy may vary depending on the complexity of the disease or institution [50]. Second, risk adjustment may be inadequate due to limited information on the case's front page and a lack of clinical details, such as the disease's severity and death information. Laboratory tests for atmospheric environmental indicators are missing. Recent studies have suggested that neutrophil percentage, red blood cell distribution width, alkaline phosphatase [51], and ambient particulate matter pollution of different sizes (PM_1, $PM_{2.5}$, and PM_{10}) [37] influence the readmission of patients with stroke prognosis after discharge. Third, the data came from only 1 hospital, and the findings may not represent other geographical areas or medical institutions, limiting the universality of the results. However, the results of the preliminary screening of factors in this study align with previous research results in different regions. Finally, our study primarily focuses on outcomes during hospitalization and mid- to long-term outcomes. However, it does not include long-term follow-up data after discharge, which limits a comprehensive assessment of patients' long-term functional recovery, risk of recurrence, and quality of life.

Conclusions

In conclusion, our findings suggest inconsistent rehospitalization rates for ischemic stroke disease at different times and that the same factors influence rehospitalization rates. Early prevention and treatment of influencing factors are required for the prognosis of ischemic stroke. Future studies should evaluate whether targeted interventions in populations with high-risk factors reduce rehospitalization rates.

Acknowledgments

This work was supported by the Ningxia Natural Science Foundation (grant number 2023AAC03445 and grant number 2020AAC03354) and the Key R&D Project of Ningxia Hui Autonomous Region (grant number 2021BEG03099).

Data Availability

The datasets used or analyzed during the current study are available from the corresponding author upon reasonable request.

Authors' Contributions

HM contributed to writing – original draft, and methodology; XW involved in validation and formal analysis; DP contributed to data curation; XS contributed to supervision, and writing – review & editing; WL contributed to project administration; ZL

XSL•FO
RenderX

contributed to supervision; YG contributed to writing – review & editing; XM contributed to supervision and data curation; TP contributed to supervision; PL contributed to supervision, and writing – review & editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary tables and figure.
[DOCX File, 1995 KB - i-jmr_v14i1e67522_app1.docx]

References

1. Roth GA, Abate D, Abate KH. Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1736-1788. [doi: 10.1016/S0140-6736(18)32203-7] [Medline: 30496103]
2. Katan M, Luft A. Global burden of stroke. Semin Neurol 2018 Apr;38(2):208-211. [doi: 10.1055/s-0038-1649503] [Medline: 29791947]
3. Ma LY, Chen WW, Gao RL, et al. China cardiovascular diseases report 2018: an updated summary. J Geriatr Cardiol 2020 Jan;17(1):1-8. [doi: 10.11909/j.issn.1671-5411.2020.01.001] [Medline: 32133031]
4. Wang YJ, Li ZX, Gu HQ, et al. China Stroke Statistics: an update on the 2019 report from the National Center for Healthcare Quality Management in Neurological Diseases, China National Clinical Research Center for Neurological Diseases, the Chinese Stroke Association, National Center for Chronic and Non-communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention and Institute for Global Neuroscience and Stroke Collaborations. Stroke Vasc Neurol 2022 Oct;7(5):415-450. [doi: 10.1136/svn-2021-001374] [Medline: 35443985]
5. Interpretation of Annual Report on Cardiovascular Health and Diseases in China 2019. Cardiology Discovery 2021;1(4):269-284. [doi: 10.1097/CD9.0000000000000040]
6. Writing committee of the report on cardiovascular health and diseases in China. Report on cardiovascular health and diseases in China 2021: an updated summary. Biomed Environ Sci 2022 Jul 20;35(7):573-603. [doi: 10.3967/bes2022.079] [Medline: 35945174]
7. Guterman EL, Douglas VC, Shah MP, Parsons T, Barba J, Josephson SA. National characteristics and predictors of neurologic 30-day readmissions. Neurology (ECronicon) 2016 Feb 16;86(7):669-675. [doi: 10.1212/WNL.0000000000002379] [Medline: 26791149]
8. Kilkenny MF, Longworth M, Pollack M, et al. Factors associated with 28-day hospital readmission after stroke in Australia. Stroke 2013 Aug;44(8):2260-2268. [doi: 10.1161/STROKEAHA.111.000531] [Medline: 23800558]
9. Lichtman JH, Leifheit-Limson EC, Jones SB, et al. Predictors of hospital readmission after stroke: a systematic review. Stroke 2010 Nov;41(11):2525-2533. [doi: 10.1161/STROKEAHA.110.599159] [Medline: 20930150]
10. Bjerkreim AT, Thomassen L, Waje-Andreassen U, Selvik HA, Næss H. Hospital readmission after intracerebral hemorrhage. J Stroke Cerebrovasc Dis 2016 Jan;25(1):157-162. [doi: 10.1016/j.jstrokecerebrovasdis.2015.09.009] [Medline: 26483156]
11. Bjerkreim AT, Thomassen L, Brøgger J, Waje-Andreassen U, Næss H. Causes and predictors for hospital readmission after ischemic stroke. J Stroke Cerebrovasc Dis 2015 Sep;24(9):2095-2101. [doi: 10.1016/j.jstrokecerebrovasdis.2015.05.019] [Medline: 26142260]
12. Rumalla K, Smith KA, Arnold PM, Mittal MK. Subarachnoid hemorrhage and readmissions: national rates, causes, risk factors, and outcomes in 16,001 hospitalized patients. World Neurosurg 2018 Feb;110:e100-e111. [doi: 10.1016/j.wneu.2017.10.089] [Medline: 29107164]
13. Smith MA, Liou JI, Frytak JR, Finch MD. 30-Day survival and rehospitalization for stroke patients according to physician specialty. Cerebrovasc Dis 2006;22(1):21-26. [doi: 10.1159/000092333] [Medline: 16567933]
14. Jencks SF, Williams MV, Coleman EA. Rehospitalizations among patients in the Medicare fee-for-service program. N Engl J Med 2009 Apr 2;360(14):1418-1428. [doi: 10.1056/NEJMsa0803563] [Medline: 19339721]
15. Ottenbacher KJ, Graham JE, Ottenbacher AJ, et al. Hospital readmission in persons with stroke following postacute inpatient rehabilitation. J Gerontol A Biol Sci Med Sci 2012 Aug;67(8):875-881. [doi: 10.1093/gerona/glr247] [Medline: 22389457]
16. Burke JF, Skolarus LE, Adelman EE, Reeves MJ, Brown DL. Influence of hospital-level practices on readmission after ischemic stroke. Neurology (ECronicon) 2014 Jun 17;82(24):2196-2204. [doi: 10.1212/WNL.0000000000000514] [Medline: 24838793]
17. Nouh AM, McCormick L, Modak J, Fortunato G, Staff I. High mortality among 30-day readmission after stroke: predictors and etiologies of readmission. Front Neurol 2017;8(632):632. [doi: 10.3389/fneur.2017.00632] [Medline: 29270149]
18. Bjerkreim AT, Naess H, Khanevski AN, Thomassen L, Waje-Andreassen U, Logallo N. One-year versus five-year hospital readmission after ischemic stroke and TIA. BMC Neurol 2019 Jan 29;19(1):15. [doi: 10.1186/s12883-019-1242-5] [Medline: 30696407]

XSL•FO
RenderX

19. Qiu X, Xue X, Xu R, et al. Predictors, causes and outcome of 30-day readmission among acute ischemic stroke. Neurol Res 2021 Jan;43(1):9-14. [doi: 10.1080/01616412.2020.1815954] [Medline: 32893753]
20. Loebel EM, Rojas M, Wheelwright D, Mensching C, Stein LK. High risk features contributing to 30-day readmission after acute ischemic stroke: a single center retrospective case-control study. Neurohospitalist 2022 Jan;12(1):24-30. [doi: 10.1177/19418744211027746] [Medline: 34950383]
21. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009 Jun 29;338:b2393. [doi: 10.1136/bmj.b2393] [Medline: 19564179]
22. van Buuren S, Groothuis-Oudshoorn K. Mice: multivariate Imputation by chained equations in R. J Stat Softw 2011 Dec;45(3):1-67. [doi: 10.18637/jss.v045.i03]
23. Deng Jianxin SL, Deqiang H, Rui T. Handling methods of missing data and their development trends. Stat Decis 2019:28-34. [doi: 10.13546/j.cnki.tjyjc.2019.23.005]
24. Le VH, Kha QH, Hung TNK, Le NQK. Risk score generated from CT-based radiomics signatures for overall survival prediction in non-small cell lung cancer. Cancers (Basel) 2021 Jul 19;13(14):3616. [doi: 10.3390/cancers13143616] [Medline: 34298828]
25. Zeng Y, Cao W, Wu C, et al. Survival prediction in home hospice care patients with lung cancer based on LASSO algorithm. Cancer Control 2022;29(10732748221124519):10732748221124519. [doi: 10.1177/10732748221124519] [Medline: 36039467]
26. Xu Y, Wang X, Huang Y, Ye D, Chi P. A LASSO-based survival prediction model for patients with synchronous colorectal carcinomas based on SEER. Transl Cancer Res 2022 Aug;11(8):2795-2809. [doi: 10.21037/tcr-20-1860] [Medline: 36093555]
27. Tibshirani R. The LASSO method for variable selection in the Cox model. Stat Med 1997;16(4). [doi: 10.1002/(sici)1097-0258(19970228)16:43.0.co;2-3] [Medline: 9044528]
28. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health 2010;13(2):273-277. [doi: 10.1111/j.1524-4733.2009.00671.x] [Medline: 19912596]
29. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Stürmer T. Variable selection for propensity score models. Am J Epidemiol 2006 Jun 15;163(12):1149-1156. [doi: 10.1093/aje/kwj149] [Medline: 16624967]
30. Austin PC. The use of propensity score methods with survival or time-to-event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat Med 2014 Mar 30;33(7):1242-1258. [doi: 10.1002/sim.5984] [Medline: 24122911]
31. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Br J Surg 2015 Feb;102(3):148-158. [doi: 10.1002/bjs.9736] [Medline: 25627261]
32. Sauerbrei W, Boulesteix AL, Binder H. Stability investigations of multivariable regression models derived from low- and high-dimensional data. J Biopharm Stat 2011 Nov;21(6):1206-1231. [doi: 10.1080/10543406.2011.629890] [Medline: 22023687]
33. Venables WN, Ripley BD. Modern applied statistics with S. In: Statistics & Computing 2002:704-705. [doi: 10.1007/978-0-387-21706-2]
34. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc 2020 Apr 1;27(4):621-633. [doi: 10.1093/jamia/ocz228] [Medline: 32106284]
35. Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol 2018 Dec;74(6):796-804. [doi: 10.1016/j.eururo.2018.08.038] [Medline: 30241973]
36. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. J Clin Oncol 2016 Jul 20;34(21):2534-2540. [doi: 10.1200/JCO.2015.65.5654] [Medline: 27247223]
37. Cai M, Lin X, Wang X, et al. Ambient particulate matter pollution of different sizes associated with recurrent stroke hospitalization in China: a cohort study of 1.07 million stroke patients. Sci Total Environ 2023 Jan 15;856(Pt 2):159104. [doi: 10.1016/j.scitotenv.2022.159104] [Medline: 36208745]
38. Lewsey J, Jhund PS, Gillies M, et al. Temporal trends in hospitalisation for stroke recurrence following incident hospitalisation for stroke in Scotland. BMC Med 2010 Apr 9;8:23. [doi: 10.1186/1741-7015-8-23] [Medline: 20380701]
39. Sun Y, Lee SH, Heng BH, Chin VS. 5-year survival and rehospitalization due to stroke recurrence among patients with hemorrhagic or ischemic strokes in Singapore. BMC Neurol 2013 Oct 3;13(133):24088308. [doi: 10.1186/1471-2377-13-133] [Medline: 24088308]
40. Becker C, Zumbrunn S, Beck K, et al. Interventions to improve communication at hospital discharge and rates of readmission: a systematic review and meta-analysis. JAMA Netw Open 2021 Aug 2;4(8):e2119346. [doi: 10.1001/jamanetworkopen.2021.19346] [Medline: 34448868]
41. Kelly C, Hulme C, Farragher T, Clarke G. Are differences in travel time or distance to healthcare for adults in global north countries associated with an impact on health outcomes? A systematic review. BMJ Open 2016 Nov 24;6(11):e013059. [doi: 10.1136/bmjopen-2016-013059] [Medline: 27884848]

XSL•FO
RenderX

42. Rao A, Barrow E, Vuik S, Darzi A, Aylin P. Systematic review of hospital readmissions in stroke patients. Stroke Res Treat 2016;2016:9325368. [doi: 10.1155/2016/9325368] [Medline: 27668120]
43. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May;46(3):399-424. [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
44. Lin HJ, Chang WL, Tseng MC. Readmission after stroke in a hospital-based registry: risk, etiologies, and risk factors. Neurology (ECronicon) 2011 Feb 1;76(5):438-443. [doi: 10.1212/WNL.0b013e31820a0cd8] [Medline: 21209374]
45. Sun Y, Toh M. Impact of diabetes mellitus (DM) on the health-care utilization and clinical outcomes of patients with stroke in Singapore. Value Health 2009;12 Suppl 3:S101-S105. [doi: 10.1111/j.1524-4733.2009.00639.x] [Medline: 20586969]
46. Gentile NT, Vaidyula VR, Kanamalla U, DeAngelis M, Gaughan J, Rao AK. Factor VIIa and tissue factor procoagulant activity in diabetes mellitus after acute ischemic stroke: impact of hyperglycemia. Thromb Haemost 2007 Nov;98(5):1007-1013. [doi: 10.1160/th06-12-0719] [Medline: 18000605]
47. Ribo M, Molina C, Montaner J, et al. Acute hyperglycemia state is associated with lower tPA-induced recanalization rates in stroke patients. Stroke 2005 Aug;36(8):1705-1709. [doi: 10.1161/01.STR.0000173161.05453.90.9f] [Medline: 16002761]
48. Khan MA, Schultz S, Othman A, et al. Hyperglycemia in stroke impairs polarization of monocytes/macrophages to a protective noninflammatory cell type. J Neurosci 2016 Sep 7;36(36):9313-9325. [doi: 10.1523/JNEUROSCI.0473-16.2016] [Medline: 27605608]
49. Huang J, Liu B, Yang C, Chen H, Eunice D, Yuan Z. Acute hyperglycemia worsens ischemic stroke-induced brain damage via high mobility group box-1 in rats. Brain Res 2013 Oct;1535(148-55):148-155. [doi: 10.1016/j.brainres.2013.08.057]
50. van Walraven C, Austin P. Administrative database research has unique characteristics that can risk biased results. J Clin Epidemiol 2012 Feb;65(2):126-131. [doi: 10.1016/j.jclinepi.2011.08.002] [Medline: 22075111]
51. Xu Y, Yang X, Huang H, et al. Extreme gradient boosting model has a better performance in predicting the risk of 90-day readmissions in patients with ischaemic stroke. J Stroke Cerebrovasc Dis 2019 Dec;28(12):104441. [doi: 10.1016/j.jstrokecerebrovasdis.2019.104441] [Medline: 31627995]

Abbreviations

AIC: Akaike information criterion
CHD: coronary heart disease
CVD: cerebrovascular disease
ICD: International Classification of Diseases
LASSO: Least Absolute Shrinkage and Selection Operator
LOS: length of stay
NIHSS: NIH Stroke Scale
SIPTW: stabilized inverse probability of treatment weighting
SMD: standardized difference

Edited by Z Guo; submitted 14.10.24; peer-reviewed by H Nzwalo, XU Ronghua; revised version received 08.04.25; accepted 01.05.25; published 03.07.25.

Please cite as:
Meng H, Wang X, Pan D, Su X, Lu W, Liu Z, Geng Y, Ma X, Pan T, Liang P
Readmission After Ischemic Stroke in Ningxia, China, From 2017 to 2021: Retrospective Cohort Study
Interact J Med Res 2025;14:e67522
URL: https://www.i-jmr.org/2025/1/e67522
doi:10.2196/67522

© Hua Meng, Xingtian Wang, Dongfeng Pan, Xinya Su, Wenwen Lu, Zhuo Liu, Yuhui Geng, Xiaojuan Ma, Ting Pan, Peifeng Liang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 3.7.2025. 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

Exploring Barriers to Patients' Progression in the Cardiac Rehabilitation Journey From Health Care Providers' Perspectives: Qualitative Study

Shri Harini Ramesh[1], MAS; Darwin Jull[1], BEng; Hélène Fournier[2], PhD; Fateme Rajabiyazdi[1,3], PEng, MSc, PhD

[1]Department of Systems and Computer Engineering, Faculty of Engineering and Design, Carleton University, Ottawa, ON, Canada

[2]National Research Council Canada, Moncton, NB, Canada

[3]Bruyère Research Institute, Bruyère, Ottawa, ON, Canada

Corresponding Author:
Fateme Rajabiyazdi, PEng, MSc, PhD
Department of Systems and Computer Engineering
Faculty of Engineering and Design
Carleton University
1125 Colonel By Dr
Ottawa, ON, K1S 5B6
Canada
Phone: 1 613 520 2600 ext 8229
Email: fateme.rajabiyazdi@carleton.ca

Abstract

Background: Cardiovascular diseases are one of the leading causes of mortality globally. Cardiac rehabilitation (CR) programs are crucial for patients recovering from cardiac events, as they help reduce the risk of recurrent events and support patient recovery. The patient's journey in CR spans the stages before, during, and after the program. Patients have to progress through each stage of CR programs successfully to complete the entire CR journey and get the full benefits of CR programs, but numerous barriers within this journey can hinder patient progression.

Objective: This study aims to explore the barriers to progression at all stages of the CR patient journey from the perspectives of health care providers involved in CR care.

Methods: This qualitative study involved semistructured interviews with health care providers involved in CR care from July 2023 to January 2024. A purposive maximal variation sampling method was used to target providers with diverse demographics and specialties. Snowball sampling was used to recruit participants, leveraging the existing networks of participants. Each interview lasted between 30 and 45 minutes. Interviews were recorded, transcribed verbatim, and analyzed using an inductive thematic analysis approach. Data analysis was conducted from August 2023 to February 2024.

Results: Ten health care providers, comprising 7 females and 3 males, were interviewed. Their roles included physician, program director, nurse manager, clinical manager, nurse coordinator, nurse, physiotherapist, and kinesiologist. The analysis identified four overarching themes related to barriers to progression in the CR journey: (1) patients not being referred to CR programs, (2) patients not enrolling in CR programs, (3) patients dropping out of CR programs, and (4) patients' lack of adherence to lifestyle changes post-CR programs.

Conclusions: In light of the growing interest in technological interventions in CR programs, we proposed 4 potential technological solutions to address the barriers to progression identified in our analysis. These solutions aim to provide a foundation for future research to guide the development of effective technologies and enhance patient progression within the CR journey.

(Interact J Med Res 2025;14:e66164) doi:10.2196/66164

KEYWORDS

cardiac rehabilitation; health care providers; CR patient journey; qualitative study; barriers; technology

XSL•FO
RenderX

Introduction

Cardiovascular diseases are one of the leading causes of mortality globally [1]. Cardiac rehabilitation (CR) programs are often recommended for patients who have experienced a cardiac event to reduce the risk of recurrent cardiovascular events and support patient recovery [2]. CR programs are comprehensive and multidisciplinary, encompassing exercise training, risk factor management, health education, and psychosocial support [3,4]. Research has shown that participation in CR programs significantly reduces morbidity and mortality associated with cardiac events [5,6] while also enhancing patients' health-related quality of life and psychological well-being [7,8].

The patient's CR journey begins even before enrollment and continues through maintaining lifestyle changes after completing the program [9,10]. It begins with a provider's referral to the CR program, made before discharge from the hospital following a cardiac event or surgery [11]. The journey then involves enrolling in, participating in, and completing the CR program [12]. After completing the program, the journey continues with integrating and maintaining the lifestyle changes and health habits learned throughout the program [13]. Patients must successfully progress through each stage to complete the entire CR journey in order to get the full benefits of CR programs. However, numerous barriers exist within the CR patient journey that could affect progression.

Previous research has identified various barriers in CR, such as low patient interest, comorbidities, travel complications, limited program availability, scheduling conflicts, and financial constraints [14-18]. However, these studies focused only on specific stages of the CR patient journey, such as referral, enrollment, completion, or postcompletion, but not throughout the entire CR patient journey. Understanding the barriers to progression throughout the entire CR patient journey is essential because each stage of the journey is interconnected, and barriers at one stage can influence subsequent stages [19,20]. Moreover, existing studies predominantly focus on the patient's perspective. It is equally important to consider the perspectives of the providers who deliver CR care, as they encounter a diverse range of patients with varying needs. Insights from those who deliver care can provide valuable information on improving CR delivery and addressing barriers effectively [21].

The aim of this study was to explore the barriers to progression throughout the CR patient journey from the perspectives of health care providers involved in CR care. We conducted semistructured interviews with 10 providers and identified 4 barriers to progression within the CR patient journey. In light of the growing interest in technological interventions in CR programs [20,22], we proposed 4 potential technological solutions to overcome the barriers to progression identified in our study. The novelty of our work lies in the identification of barriers to progression from the perspectives of health care providers and the proposal of potential solutions to address these challenges. We hope these potential solutions provide the foundation for future research and guide the development of effective technologies to overcome barriers to progression within the CR patient journey.

Methods

Ethical Considerations

This study protocol was approved by the ethics committees of Carleton University (clearance number 119779) and the National Research Council Canada (clearance number 2023-108). Ethical review was conducted per institutional guidelines, with no exemptions or waivers applied. Written informed consent was obtained before each interview. Participants were informed of their right to withdraw at any time without consequence. No secondary data were used. All data were anonymized to ensure privacy and confidentiality. No compensation was provided to the participants.

Overview

This qualitative study is reported in accordance with the COREQ (COnsolidated criteria for REporting Qualitative research) guidelines [23] (Multimedia Appendix 1). We conducted semistructured interviews to explore the barriers to progression in the CR patient journey from the perspective of the health care providers involved in CR care. We chose interviews to ensure that participants could freely share their thoughts and experiences. The semistructured format allowed us to ask follow-up questions in order to gain an in-depth understanding of their perspectives. In addition, the confidential nature of the interviews encouraged open and honest sharing [24]. Eligible participants included health care providers who refer patients to CR, those who provide CR services, and those who manage CR programs. Eligible participants were also required to be proficient in English. Exclusion criteria included individuals not involved in direct patient care or CR, such as administrative staff. To ensure that the perspectives of a diverse group of participants from various specialties in CR care were included, we used purposive sampling [25]. To further expand the participant pool, we used snowball sampling methods [26]. HF, who has extensive connections with health care providers involved in CR care and CR centers, shared the study objectives and recruitment details within her network. Through this process, she received the email addresses of potential participants. Subsequently, SHR contacted these potential participants via email, providing detailed information about the study and inviting them to participate. We contacted 11 potential participants, of whom 10 agreed to participate, while 1 did not respond. No participants dropped out after providing informed consent. No prior relationship was established between the researcher and the participants before the study commenced, and participants knew only that the researcher was conducting the study as part of an academic project focused on CR.

Data Collection

Data collection took place from July 2023 to January 2024. The interviews were conducted virtually via Microsoft Teams, with each session lasting between 30 and 45 minutes. No repeat interviews were carried out. Each participant was interviewed once during the study. The researcher (SHR), a female PhD student familiar with the subject matter and trained in qualitative

XSL•FO
RenderX

research, conducted the interviews, asked questions and follow-up questions as needed, and took notes during the sessions. While conducting the interviews, the researcher maintained a neutral stance, ensuring that her personal views or assumptions did not influence the questions asked or the participants' responses. The research team collaboratively and iteratively developed an interview script and questions (Multimedia Appendix 2). The questions were open-ended to allow for an in-depth exploration of the providers' perspectives on the barriers to progression in the CR patient journey, with follow-up questions asked when appropriate. We conducted a pilot test of the interview guide with a health researcher to ensure clarity, relevance, and logical flow of the questions. Feedback from this pilot test was used to refine the questions and the interview guide before data collection commenced.

Data Analysis

Our analysis was conducted from August 2023 to February 2024 inductively and did not aim to validate any preexisting theories or hypotheses. We followed a thematic analysis approach following the 6 steps outlined by Braun and Clarke [27]. Two analysts, SHR and a male undergraduate student (DJ) with subject knowledge and qualitative research training, were responsible for the analysis. Disagreements in coding and theme assignment were resolved through consensus facilitated by senior female researchers FR and HF, both of whom are familiar with the subject matter and hold PhDs with extensive experience in qualitative research. We adopted a reflexive approach, acknowledging the influence of researchers' views and experiences [28]. In step 1, both analysts read interview transcripts and took notes. In step 2, they reviewed the first 2 transcripts line by line and identified prominent codes to generate a codebook using NVivo software (version 14.23; Lumivero), which was then applied to subsequent interviews. Codes were iteratively refined every 2 interviews. Step 3 involved systematically grouping codes into candidate themes. In step 4, these themes were compared and refined. In step 5, other team members contributed to further theme refinement. Thematic saturation was assessed using a saturation table that documented the themes found in each interview. Finally, in step 6, we produced a report summarizing our findings. Transcripts were not returned to participants for comment or correction. The analysis was conducted based solely on the interview data collected during the sessions.

Rigor and Trustworthiness

Rigor and trustworthiness were achieved by ensuring credibility, transferability, dependability, and confirmability throughout the research process [29]. To address credibility, we conducted peer debriefing sessions within our research team, which provided an external check on our interpretation of the data, analysis, and coding decisions. Dependability was achieved through detailed documentation of themes and subthemes, providing transparency, and enabling replicability of the coding process. To support transferability, we provided comprehensive descriptions of our study design, participant selection criteria, and data collection methods, allowing others to evaluate the applicability of our findings to various CR settings. As Lincoln and Guba [30] highlighted, confirmability is achieved when credibility, transferability, and dependability are established. We ensured confirmability by documenting the rationale for theoretical, methodological, and analytical decisions throughout the study.

Results

Overview

Ten health care providers, comprising 7 females and 3 males involved in CR care, including a physician, a program director, a clinical manager, a nurse manager, a nurse coordinator, a nurse, 3 physiotherapists, and a kinesiologist, were interviewed (Table 1). The thematic analysis revealed four overarching themes concerning barriers to progression within the CR patient journey, as perceived by these providers: (1) patients not being referred to CR programs, (2) patients not enrolling in CR programs, (3) patients dropping out of CR programs, and (4) patients' lack of adherence to lifestyle changes post-CR programs. Multimedia Appendix 3 provides illustrative quotes within each of the themes identified. Thematic saturation was reached after analyzing 7 interviews [31] (Multimedia Appendix 4). This was determined using a saturation table, where no new themes emerged after the seventh interview. To ensure robustness, 3 additional interviews were conducted, which confirmed the saturation of themes. The early achievement of saturation reflects the shared perspectives of participants, despite their varied roles within CR care.

XSL•FO
RenderX

Table 1. Characteristics of interviewed participants.

Characteristics	Values
Total, n	10
Sex, n (%)	
Female	7 (70)
Male	3 (30)
Specialty, n (%)	
Clinical manager	1 (10)
Nurse manager	1 (10)
Nurse	1 (10)
Physician	1 (10)
CR[a] program director	1 (10)
Kinesiologist	1 (10)
Physiotherapist	3 (10)
Nurse coordinator	1 (10)
Practice setting, n (%)	
Urban	7 (70)
Urban and rural	2 (20)
Rural	1 (10)
Age (year), mean (SD)	47.9 (10.2)
Years of experience, mean (SD)	18.7 (9.5)

[a]CR: cardiac rehabilitation.

Patients Not Being Referred to CR Programs

One of the barriers to progression within the CR patient journey is patients not being referred to CR programs, as without referrals, patients may not have access to CR programs (Table S1 in Multimedia Appendix 5). Unintentional biases in referrals, such as those related to health conditions, geography, age, and gender, can contribute to patients not being referred to CR programs. As one provider noted, “There is a lot of inherent bias in terms of who is being referred to rehab” [P3]. Providers may hesitate to refer patients with comorbidities or complex health conditions to CR programs due to concerns about their ability to participate. In addition, if the patient is not residing near a CR center, the provider may assume that it is difficult for the patient to travel and, as a result, be hesitant to refer the patient to a CR program. Providers may perceive that older adults might be frail or disconcerted after a cardiac event or surgery, which may result in providers not referring these patients to a CR program. Some providers may be hesitant to refer certain patients to CR programs, considering their gender-specific responsibilities, such as work, family, and parental roles, which could contribute to women being overlooked for referrals.

Furthermore, another reason for not referring is the providers’ heavy workloads. The paperwork involved in referring patients to CR programs can be time-consuming, which may discourage some providers from making these referrals. Although automatic referral systems are available in some CR centers, providers’ endorsements are crucial for motivating patients; as one provider emphasized, “Endorsements from [providers] are key for [motivating patients]” [P1]. However, individually discussing the benefits of CR with every patient is time-consuming and requires much effort. Referring providers may also lack awareness of the various programs available at the CR center. As noted, “[Referring providers] just don’t know all the services that are available” [P8]. Consequently, providers may hesitate to refer patients due to uncertainty about the programs available. In addition, providers mentioned that it is important to carefully select a CR program that aligns with the patient’s needs, as some programs with high-impact exercises might be appropriate for immediate needs, whereas low-impact exercise-based programs could offer better long-term benefits. However, identifying the best CR program for each patient during referral can be challenging for providers without access to a wholesome list of CR programs.

Patients Not Enrolling in CR Programs

Another barrier to progression in the CR patient journey is patients not enrolling in CR programs after referral. After patients are referred, they must enroll and start the program to benefit from it. Providers mentioned various factors contributing to this nonenrollment (Table S2 in Multimedia Appendix 5). Providers often observed that a substantial number of patients lack awareness and knowledge about CR, with many patients being unaware of its importance and benefits. Providers also noted that patients sometimes hold misconceptions about CR, believing that they can recover by themselves at home or that

XSL•FO
RenderX

surgical treatment alone is sufficient without rehabilitation. Providers think that inconvenient waiting periods also contribute to patients not enrolling in CR programs. For some patients, the waiting period may be perceived as too short after surgery, leading them to feel unready to start the program. Conversely, for other patients, the waiting period between referral and the program's start may seem excessively long. In addition, CR programs are sometimes offered on fixed schedules, where groups of patients begin the program simultaneously. If patients miss the upcoming enrollment window, they will need to wait for the next scheduled start date, which can lead to delays in enrolling in CR programs.

Providers also think that cultural restrictions can prevent some patients from enrolling in CR programs, particularly when there is discomfort or disallowance in mixed-gender sessions due to cultural or religious beliefs. Providers have mentioned that financial barriers are a concern for some patients and can affect their enrollment. Some patients may not have insurance coverage, or they may be unaware of what services their insurance will cover. Financial barriers can also include the costs associated with transportation and parking. In addition, in some CR centers, patients have the option to participate in CR programs virtually through videoconferencing and web-based meeting platforms such as Zoom (Zoom Video Communications). Providers mentioned a lack of technical knowledge and equipment requirements as factors contributing to nonenrollment in these virtual CR (VCR) programs. They noted that patients may encounter obstacles due to the need for reliable Wi-Fi, laptops, tablets, email addresses, and smartwatches to monitor vitals such as heart rate for VCR, which may not be accessible to all patients.

While providers have mentioned several factors for patients not enrolling in CR programs, they believe that some factors remain hidden and express uncertainty about fully understanding why some patients do not enroll in these programs. Providers have also highlighted that while they try to find these factors through patient data recorded in electronic medical records (EMRs), they are unable to fully explore and uncover the underlying reasons solely from EMRs due to missing or insufficient data. As one provider noted, "We try and capture what we can from [the EMR], so we can count visits and things like that within the EMR. It is definitely not perfect" [P7]. Providers expressed the need for a tracking system to understand potential hidden factors that contribute to patients not enrolling. As one provider mentioned, "[We] do need a little bit of like...something to keep track of, because there's quite a few patients that come through. So just to have...what date they start...it is nice to have an overview of the patient's journey" [P8].

Patients Dropping Out of CR Programs

Dropping out of a CR program before completing it is also a barrier to progression in the CR patient journey, as it prevents patients from fully benefiting from the CR program (Table S3 in Multimedia Appendix 5). Providers observed that reproductive and hormonal conditions can deter female patients from completing or fully participating in CR programs. These conditions include concerns about the safety of exercise during pregnancy, managing pregnancy-related fatigue, menopause issues, and menstrual cycle discomfort, such as severe menstrual cramps. In addition, providers recognized that a lack of support is a factor for dropout, including insufficient family support to manage home obligations such as caregiving for children or older adult family members or a lack of peer support for mental and emotional well-being during the rehabilitation process.

Providers noted that low motivation, low self-efficacy, and a lack of belief in one's ability to perform exercises or improve one's lifestyle contribute to a lack of participation and eventual dropout from the program. As one provider mentioned, "They just don't [participate] because they have no motivation to" [P8]. In addition, providers have observed that low accountability in some patients, requiring an external push and motivation from family or providers, can also lead to dropout. Although providers make efforts to encourage patients by emphasizing the benefits of continuing rehabilitation, some patients may still find it challenging to stay engaged with their rehabilitation, which can contribute to dropouts. As one provider mentioned, "We try to help them understand that if they [continue rehabilitation], their health can get better" [P9].

Patients may face several challenges in reaching CR centers, particularly when they lack access to a car or public transportation or when adverse weather conditions, such as extreme heat or cold, are present. Providers mentioned that these challenges in reaching CR centers can contribute to patient dropouts. In addition, the lack of multilingual support and translators in some CR programs can be challenging for patients who need services in other languages. Providers mentioned that these language barriers can also contribute to dropouts. Providers highlight that some patients may be frail due to medical conditions such as shortness of breath, arthritis, or chronic pain, which can prevent them from performing the required exercises, leading to frustration and dropout. Furthermore, in VCR, patients often need to interact with digital interventions for extended periods. Providers mentioned that this prolonged interaction can be one of the factors contributing to dropouts, particularly in VCR.

Patients' Lack of Adherence to Lifestyle Changes Post-CR Programs

Providers recommend that patients continue following the lifestyle changes they learned during the CR program even after finishing the CR program cycle to achieve better health-related quality of life. However, patients sometimes struggle to adhere to these recommendations, which becomes a barrier to progression in their journey (Table S4 in Multimedia Appendix 5). Providers noted that patients could experience persistent feelings of sadness, hopelessness, stress, and anxiety about daily life and responsibilities, which could lower their motivation and hinder their adherence to lifestyle changes. A lack of personal drive could also contribute to not adhering, as patients may struggle to set and achieve personal health goals independently. Providers noted that financial constraints may impact adherence to post-CR programs, with high costs for healthy food, gym memberships, medications, and inadequate insurance coverage preventing patients from fully following recommended lifestyle changes. Providers try to emphasize the importance of physical exercise in maintaining a healthy

XSL•FO
RenderX

lifestyle, but only a small portion of patients follow this advice. As one provider noted, "The reality is that a small percentage of patients, once they graduate rehab, remain as physically active as they were during rehab" [P3]. In addition, providers mentioned that some patients face challenges in performing physical exercises due to factors such as lack of space at home, exercise-induced pain or discomfort, and a lack of resources to fully understand and perform exercises without pain or discomfort. These challenges lead to frustration and doubts about their ability to perform exercises, as mentioned by a provider: "I do get a few [patients] that tell me they're just not interested because exercise is just not for them" [P9]. Providers think that a lack of monitoring and follow-up clinical appointments could also contribute to nonadherence, with insufficient regular feedback on health progress and areas for improvement leading to decreased adherence to lifestyle changes.

Discussion

Technological Solutions

In this study, we explored the barriers to progression in the CR patient journey from the health care providers' perspectives. We identified four barriers that hinder progression in the CR patient journey, such as (1) patients not being referred to CR programs, (2) patients not enrolling in CR programs, (3) patients dropping out of CR programs, and (4) patients' lack of adherence to lifestyle changes post-CR programs. Based on our findings, we recommend 4 potential technological solutions to overcome these barriers: monitoring patients' progress throughout the CR journey using interactive data visualization systems, monitoring physical exercises through automated video analysis–based feedback systems, matching the patients to the right programs using machine learning (ML)–driven predictive analysis, and supporting patients with self-management using targeted natural language processing (NLP) and large language model (LLM) tools. These potential solutions could provide the foundation for future research and guide the development of effective technologies to overcome barriers to progression in the CR patient journey.

Monitoring Patients' Progress Throughout the CR Journey Using Interactive Data Visualization Systems

Health care providers have noticed that some patients in CR programs do not show adequate improvement in their progress, which could be due to the loss of motivation over time. In addition, they have mentioned that there may be other hidden reasons within the CR programs that are impacting patients' progress, making it difficult for patients to take necessary actions to improve progression. Effective monitoring of the patient's progress can help patients stay motivated [32-34] and help providers identify any hidden reasons affecting patients' progress [35,36]. Research has shown that interactive data visualization systems are effective in monitoring patients' progress for both patients [37,38] and providers [39,40]. Based on the success of these systems, we propose the development of interactive data visualization tools for monitoring patients' progress in CR. Patients and providers have different goals and require varying levels of detail in monitoring patients' progress [41]. Therefore, visualization systems should be developed targeting the unique needs of patients and providers with a human-centered approach [42] while involving them in the process of design and development.

For patients, visualization systems can enable them to monitor their progress throughout the CR journey by using EMR and personally collected health data in CR, such as vital signs, exercise performance, attendance, and assessment results. The system can include gamified visualizations [43] and real-life object representations [44] to make it engaging and easier for patients to understand their progress, rather than using complex charts. For example, a patient who visually sees an improvement in heart rate rhythm and blood pressure after regular exercise may feel more encouraged to actively participate throughout the CR journey.

For providers, visualization systems can assist them in monitoring the progress of patients in CR journey and help them identify any underlying reasons that affect patient progression by using aggregated EMR data of patients in the CR center. Providers can use the visualization system to filter and drill down into specific patients or groups based on demographic characteristics, particular cohorts, or specific CR sessions to analyze patients' attendance patterns and improvements or declines in progress [45]. For instance, providers might discover a trend where patients older than 65 years show better progress in heart rate rhythm when exercise sessions are coupled with meditation than when they are offered solely.

Monitoring Physical Exercises Using Automated Video Analysis–Based Feedback Systems

Providers noticed that many patients in CR struggle to perform physical exercises correctly, particularly during VCR and after completing CR programs, where direct supervision from health care providers is not available. Without this supervision and real-time feedback, patients often find it challenging to understand proper exercise techniques, leading to uncertainty about their form and an increased risk of adopting incorrect postures or unsafe practices, which could result in injuries. Recent studies show promising results for monitoring physical exercises with artificial intelligence–powered feedback systems that analyze posture [46,47] and provide real-time exercise feedback through video analysis [48-50]. Building on the success of previous artificial intelligence–powered feedback systems, we suggest developing automated video analysis–based physical exercise feedback systems to guide and monitor patients' exercise techniques and provide immediate corrective feedback.

The system can leverage pose estimation algorithms such as DensePose [51], YOLO-Pose [52], and OpenPose [53] to identify patient posture by detecting joints in real-time video. These algorithms can be trained on a dataset including correct and incorrect postures, as demonstrated and annotated by health care providers [54]. The system can then incorporate mathematical algorithms that calculate the angles between various body joints to classify whether a posture is correct or incorrect [47]. Patients can perform physical exercises while the system captures video through their web cameras. The system would analyze the video feed in real time, classifying the exercise form as correct or incorrect. If an incorrect form is

XSL•FO
RenderX

detected, the system will identify the specific area of the body that is not aligned properly and provide immediate feedback. This feedback could be delivered through text and image overlays, offering suggestions such as "You are doing a great job, but bend more," alongside visual corrections that highlight the areas needing adjustment. In addition, real-time voice prompts could be integrated to help patients quickly understand and adjust their form during the exercise. For example, if a patient is performing a lunge and their front knee extends too far forward over their toes, the system might display a red overlay on the knee area with a voice prompt saying, "Step back slightly to keep your knee aligned with your ankle."

Matching the Patients to the Right Programs Using ML-Driven Predictive Analysis

Health care providers mentioned that patients often drop out early from CR programs, which could be because they are not matched to a program that meets their needs during referral [55,56]. Providers face challenges in identifying which patients are at risk of dropping out and selecting the appropriate program that aligns with those needs, as this process is complex and time-consuming. Previous research has demonstrated the efficacy of predictive algorithms in successfully predicting dropout rates and selecting appropriate programs. For instance, predictive algorithms have been used to predict dropout rates in health care programs [57,58], such as cognitive behavioral therapy for insomnia [59], chronic disease management programs [60], rehabilitation programs [61-63], and for selecting appropriate treatment plans [64,65]. Building on previous research, we suggest that ML-driven predictive analysis algorithms can be developed to proactively identify patients at heightened risk of dropping out from CR programs based on their characteristics and historical program outcomes and suggest the most suitable CR programs for each patient.

These algorithms can be developed using various ML models. For example, logistic regression [66] or naive Bayes models [67] can assess how specific patient factors influence the likelihood of dropout. Random forests [68], or eXtreme Gradient Boosting models [69], can be used to identify which factors (or combinations of factors) are most predictive of dropouts. Gradient boosting machines [70] or support vector machines [71] can then predict dropout risk and recommend the most suitable CR programs based on the identified factors. These algorithms can be trained using historical patient data such as age, gender, location, marital status, education level, socioeconomic status, comorbidities, and previous CR program and health outcomes (eg, completion, dropout, improvement in cardiac health, readmission, or relapse). Once trained, the algorithm can be used to match new patients to the right CR program at the point of referral. By comparing the characteristics of new patients with those of previous patients who dropped out, the algorithms can pinpoint factors associated with dropout risks, such as irregular attendance or poor performance in rehabilitation sessions. The algorithm can generate risk scores that indicate the new patients' likelihood of dropout. It can then recommend the CR program that has historically yielded the best outcomes for patients with similar profiles. For instance, a 65-year-old patient with a history of diabetes and hypertension and a 90% dropout risk due to long travel distance might be matched with a program offering fewer inpatient visits or a virtual option, emphasizing dietary management and low-impact exercises based on the success of similar profiles in historical data.

Supporting Patients With Self-Management Using Targeted NLP and LLM Tools

Providers observed that many patients in CR face difficulties in self-management. This includes difficulties in independently setting their health goals and finding reliable answers to their health-related questions outside of their scheduled appointments with providers. To help patients set their health goals on their own, NLP-based automated goal-setting tools have been shown to be effective in various medical conditions such as hypertension [72], HIV [73], stroke [74], and aphasia [75]. Similarly, to help patients find answers to their health-related questions, LLM-based chatbot solutions have been shown to be successful, particularly, for patients managing chronic diseases [76], inflammatory bowel diseases [77], and breast cancer [78]. Building on the effectiveness of these tools in various medical conditions, we suggest developing targeted NLP- and LLM-based tools specifically for patients in CR. NLP-based tools, such as mobile apps or web interfaces, can assist patients in setting personal health goals in CR. LLM-integrated chatbots can help patients find answers to their questions and enhance their understanding throughout their CR journey.

For goal setting, NLP-based tools can allow patients to input detailed information about their health conditions, preferences, goals, and obstacles in their own words. Using NLP techniques, the tool can use intent recognition [79] to identify patient objectives, entity extraction [80,81] to categorize health conditions and challenges, and sentiment analysis [82,83] to assess emotional states and motivation levels. Based on this analysis, the technology can generate personalized goals using the SMART (Specific, Measurable, Achievable, Relevant, Time-bound) framework [84]. Patients can also review, refine, prioritize, and track their goals over time. For example, if a patient wants to "exercise more" but struggles with "low energy levels," the technology could suggest a goal based on the SMART framework, such as "Walk for 20 minutes, three times a week, and gradually increase the duration as energy improves."

To provide answers to patients' questions, targeted LLM-based chatbots can be trained on CR-specific educational resources, a database of frequently asked questions by patients in CR, and anonymized patient-provider conversations specific to CR. The chatbot should be capable of providing CR education, detailed explanations to patients' questions; offering empathetic responses and coping strategies for managing stress, anxiety, and other emotional challenges; and assisting patients in finding support groups and educational materials tailored to their needs. If patients have questions anytime throughout the CR patient journey, they can access the chatbot. For instance, if a patient asks how to safely continue their exercise routine at home while experiencing menstrual cramps, the chatbot could recommend gentle stretching exercises or advise reducing workout intensity while encouraging confirmation of this guidance with their provider at their next appointment [85].

XSL•FO
RenderX

Limitations

While our study aimed to capture the perspectives of health care providers from various specialties involved in CR care, the use of snowball sampling may have introduced selection bias, as participants were primarily recruited based on a research collaboration network. Although we reached thematic saturation after 7 interviews, indicating common barriers across providers from different specialties, the small sample size of 10 participants may not fully reflect the diversity of perspectives, particularly as we did not include providers from other countries. In addition, despite following rigorous methodological standards, the qualitative analysis remains interpretative by nature and may be subject to researcher bias [84]. As with any qualitative study, the rigor of our findings should be judged by their resonance and plausibility rather than their generalizability. Furthermore, our findings may not be transferable to other CR centers that differ in resources, cultures, or geographical settings [85].

Conclusions

The objective of this study was to explore the barriers to progression in the CR patient journey from the perspectives of health care providers involved in CR care. Our findings show that patients not being referred to CR programs, patients not enrolling in CR programs, patients dropping out of CR programs, and patients' lack of adherence to lifestyle changes post-CR programs are the 4 barriers to progression as perceived by providers. We also proposed 4 potential technological solutions to overcome the barriers identified through our study. Future work should focus on designing, developing, and evaluating the technological solutions proposed in this study to overcome the barriers to progression in the CR patient journey.

Acknowledgments

This study was funded by the National Research Council Aging in Place program and the Natural Sciences and Engineering Research Council (NSERC) Discovery Grant. The authors are grateful to all the participants for sharing their invaluable perspectives, without whom this study would not have been possible.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (COnsolidated criteria for REporting Qualitative research) checklist.
[PDF File (Adobe PDF File), 74 KB - ijmr_v14i1e66164_app1.pdf]

Multimedia Appendix 2
Semistructured interview script and questions.
[PDF File (Adobe PDF File), 73 KB - ijmr_v14i1e66164_app2.pdf]

Multimedia Appendix 3
Representative quotes on barriers to progression in the cardiac rehabilitation patient journey.
[PDF File (Adobe PDF File), 128 KB - ijmr_v14i1e66164_app3.pdf]

Multimedia Appendix 4
Saturation table.
[PDF File (Adobe PDF File), 135 KB - ijmr_v14i1e66164_app4.pdf]

Multimedia Appendix 5
Codebook.
[PDF File (Adobe PDF File), 81 KB - ijmr_v14i1e66164_app5.pdf]

References

1. Cardiovascular diseases (CVDs). World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) [accessed 2024-08-15]
2. Taylor RS, Dalal HM, McDonagh STJ. The role of cardiac rehabilitation in improving cardiovascular outcomes. Nat Rev Cardiol 2022;19(3):180-194 [FREE Full text] [doi: 10.1038/s41569-021-00611-7] [Medline: 34531576]
3. Balady GJ, Williams MA, Ades PA, Bittner V, Comoss P, Foody JM, American Heart Association Exercise, Cardiac Rehabilitation,Prevention Committee, the Council on Clinical Cardiology, American Heart Association Council on Cardiovascular Nursing, American Heart Association Council on EpidemiologyPrevention, American Heart Association Council on Nutrition, Physical Activity,Metabolism, American Association of CardiovascularPulmonary Rehabilitation. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the

XSL•FO
RenderX

American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2007;115(20):2675-2682. [doi: 10.1161/CIRCULATIONAHA.106.180945] [Medline: 17513578]
4. Protoview P, Vol B, Ringgold B. Guidelines for Cardiac Rehabilitation Programs, 6th Edition. Champaign, IL: Human Kinetics; 2020.
5. Long L, Mordi I, Bridges C, Sagar VA, Davies EJ, Coats AJS, et al. Exercise-based cardiac rehabilitation for adults with heart failure. Cochrane Database Syst Rev 2019;1(1):CD003331 [FREE Full text] [doi: 10.1002/14651858.CD003331.pub5] [Medline: 30695817]
6. Anderson L, Thompson DR, Oldridge N, Zwisler A, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease. Cochrane Database Syst Rev 2016;2016(1):CD001800 [FREE Full text] [doi: 10.1002/14651858.CD001800.pub3] [Medline: 26730878]
7. Jiménez-Navarro MF, Lopez-Jimenez F, Pérez-Belmonte LM, Lennon RJ, Diaz-Melean C, Rodriguez-Escudero JP, et al. Benefits of cardiac rehabilitation on cardiovascular outcomes in patients with diabetes mellitus after percutaneous coronary intervention. J Am Heart Assoc 2017;6(10):e006404 [FREE Full text] [doi: 10.1161/JAHA.117.006404] [Medline: 29021270]
8. Yohannes AM, Doherty P, Bundy C, Yalfani A. The long-term benefits of cardiac rehabilitation on depression, anxiety, physical activity and quality of life. J Clin Nurs 2010;19(19-20):2806-2813. [doi: 10.1111/j.1365-2702.2010.03313.x] [Medline: 20738450]
9. Campbell M, McEvoy JW, Calpin G, Concannon F, Redfern S. Perspectives of healthcare professionals on clinician-patient communication of cardiovascular disease risk. J Patient Exp 2024;11:23743735241257386 [FREE Full text] [doi: 10.1177/23743735241257386] [Medline: 38807917]
10. Vanzella LM, Cotie LM, Flores-Hukom M, Marzolini S, Konidis R, Ghisi GLDM. Patients' perceptions of hybrid and virtual-only care models during the cardiac rehabilitation patient journey: a qualitative study. J Cardiovasc Nurs 2025;40(2):E91-E100. [doi: 10.1097/JCN.0000000000001076] [Medline: 38206327]
11. Jackson L, Leclerc J, Erskine Y, Linden W. Getting the most out of cardiac rehabilitation: a review of referral and adherence predictors. Heart 2005;91(1):10-14 [FREE Full text] [doi: 10.1136/hrt.2004.045559] [Medline: 15604322]
12. Ritchey MD, Maresh S, McNeely J, Shaffer T, Jackson SL, Keteyian SJ, et al. Tracking cardiac rehabilitation participation and completion among medicare beneficiaries to inform the efforts of a national initiative. Circ Cardiovasc Qual Outcomes 2020;13(1):e005902 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.119.005902] [Medline: 31931615]
13. Janssen V, De Gucht V, van Exel H, Maes S. A self-regulation lifestyle program for post-cardiac rehabilitation patients has long-term effects on exercise adherence. J Behav Med 2014;37(2):308-321. [doi: 10.1007/s10865-012-9489-y] [Medline: 23334387]
14. Platz K, Kools S, Howie-Esquivel J. Benefits, facilitators, and barriers of alternative models of cardiac rehabilitation: a qualitative systematic review. J Cardiopulm Rehabil Prev 2023;43(2):83-92. [doi: 10.1097/HCR.0000000000000738] [Medline: 36346781]
15. Sérvio TC, Britto RR, de Melo Ghisi GL, da Silva LP, Silva LDN, Lima MMO, et al. Barriers to cardiac rehabilitation delivery in a low-resource setting from the perspective of healthcare administrators, rehabilitation providers, and cardiac patients. BMC Health Serv Res 2019;19(1):615 [FREE Full text] [doi: 10.1186/s12913-019-4463-9] [Medline: 31477103]
16. Resurrección DM, Motrico E, Rigabert A, Rubio-Valera M, Conejo-Cerón S, Pastor L, et al. Barriers for nonparticipation and dropout of women in cardiac rehabilitation programs: a systematic review. J Womens Health (Larchmt) 2017;26(8):849-859. [doi: 10.1089/jwh.2016.6249] [Medline: 28388314]
17. Ozemek C, Squires RW. Enrollment and adherence to early outpatient and maintenance cardiac rehabilitation programs. J Cardiopulm Rehabil Prev 2021;41(6):367-374. [doi: 10.1097/HCR.0000000000000645] [Medline: 34727555]
18. Supervía M, Medina-Inojosa JR, Yeung C, Lopez-Jimenez F, Squires RW, Pérez-Terzic CM, et al. Cardiac rehabilitation for women: a systematic review of barriers and solutions. Mayo Clin Proc 2017;92(4):565-577 [FREE Full text] [doi: 10.1016/j.mayocp.2017.01.002] [Medline: 28365100]
19. Gravely-Witte S, Leung YW, Nariani R, Tamim H, Oh P, Chan VM, et al. Effects of cardiac rehabilitation referral strategies on referral and enrollment rates. Nat Rev Cardiol 2010;7(2):87-96 [FREE Full text] [doi: 10.1038/nrcardio.2009.223] [Medline: 19997077]
20. Ades PA, Keteyian SJ, Wright JS, Hamm LF, Lui K, Newlin K, et al. Increasing cardiac rehabilitation participation from 20% to 70%: a road map from the million hearts cardiac rehabilitation collaborative. Mayo Clin Proc 2017;92(2):234-242 [FREE Full text] [doi: 10.1016/j.mayocp.2016.10.014] [Medline: 27855953]
21. O'Doherty AF, Humphreys H, Dawkes S, Cowie A, Hinton S, Brubaker PH, et al. How has technology been used to deliver cardiac rehabilitation during the COVID-19 pandemic? An international cross-sectional survey of healthcare professionals conducted by the BACPR. BMJ Open 2021;11(4):e046051 [FREE Full text] [doi: 10.1136/bmjopen-2020-046051] [Medline: 33879492]
22. Beatty AL, Beckie TM, Dodson J, Goldstein CM, Hughes JW, Kraus WE, et al. A new era in cardiac rehabilitation delivery: research gaps, questions, strategies, and priorities. Circulation 2023;147(3):254-266 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.122.061046] [Medline: 36649394]

XSL•FO
RenderX

23. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
24. Britten N. Qualitative interviews in medical research. BMJ 1995;311(6999):251-253 [FREE Full text] [doi: 10.1136/bmj.311.6999.251] [Medline: 7627048]
25. Patton MQ. Qualitative Evaluation and Research Methods, 2nd ed. Thousand Oaks, CA: Sage Publications, Inc; 1990.
26. Biernacki P, Waldorf D. Snowball sampling: problems and techniques of chain referral sampling. Soc Methods Res 1981;10(2):141-163 [FREE Full text] [doi: 10.1177/004912418101000205]
27. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
28. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995;311(6996):42-45 [FREE Full text] [doi: 10.1136/bmj.311.6996.42] [Medline: 7613329]
29. Ahmed SK. The pillars of trustworthiness in qualitative research. J. Public Health Med 2024;2:100051. [doi: 10.1016/j.glmedi.2024.100051]
30. Lincoln YS, Guba EG. But is it rigorous? Trustworthiness and authenticity in naturalistic evaluation. New Dir Program Eval 2004;1986(30):73-84. [doi: 10.1002/ev.1427]
31. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
32. Cajamarca G, Herskovic V, Rossel PO. Enabling older adults' health self-management through self-report and visualization—a systematic literature review. Sensors (Basel) 2020;20(15):4348 [FREE Full text] [doi: 10.3390/s20154348] [Medline: 32759801]
33. Polhemus A, Novak J, Majid S, Simblett S, Morris D, Bruce S, et al. Data visualization for chronic neurological and mental health condition self-management: systematic review of user perspectives. JMIR Ment Health 2022;9(4):e25249 [FREE Full text] [doi: 10.2196/25249] [Medline: 35482368]
34. Maitland J, Chalmers M. Self-monitoring, self-awareness, and self-determination in cardiac rehabilitation. 2010 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2010 April 10; Atlanta, GA p. 1213-1222. [doi: 10.1038/s41569-021-00611-7]
35. McCarthy S, O'Raghallaigh P, Woodworth S, Lim YL, Kenny LC, Adam F. An integrated patient journey mapping tool for embedding quality in healthcare service reform. J Decis Syst 2016;25(sup1):354-368. [doi: 10.1145/1753326.1753508]
36. Borycki EM, Kushniruk AW, Wagner E, Kletke R. Patient journey mapping: integrating digital technologies into the journey. Knowl Manag e-learn Int J 2020;12(4):521-535. [doi: 10.34105/j.kmel.2020.12.029]
37. Chen L, Yan N, Kiang M, Muth A, Krishna K. Innomotion: a web-based rehabilitation system helping patients recover and gain self-awareness of their body away from the clinic. 2014 Presented at: CHI '14 Extended Abstracts on Human Factors in Computing Systems; 2014 April 26; Ontario, Toronto, Canada p. 233-238. [doi: 10.1145/2559206.2580926]
38. Zhu J, Lei Y, Shah A, Schein G, Ghaednia H, Schwab J, et al. MuscleRehab: Improving unsupervised physical rehabilitation by monitoring and visualizing muscle engagement. 2022 Presented at: UIST '22: Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology; 2022 October 28; Bend, OR p. 1-14. [doi: 10.1145/3526113.3545705]
39. Shafei I, Karnon J, Crotty M. Process mining and customer journey mapping in healthcare: enhancing patient-centred care in stroke rehabilitation. Digit Health 2024;10:20552076241249264 [FREE Full text] [doi: 10.1177/20552076241249264] [Medline: 38766357]
40. Widanagamaachchi W, Livnat Y, Bremer P, Duvall S, Pascucci V. Interactive visualization and exploration of patient progression in a hospital setting. AMIA Annu Symp Proc 2017;2017:1773-1782 [FREE Full text] [Medline: 29854248]
41. Rajabiyazdi F, Perin C, Oehlberg L, Carpendale S. Exploring the design of patient-generated data visualizations. 2020. URL: https://hal.science/hal-02861239 [accessed 2024-08-11]
42. Maguire M. Methods to support human-centred design. Int J Hum Comput Stud 2001;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
43. Feng Z, Gao Y, Zhang T. Gamification for visualization applications in the construction industry. Struct Integr 2022;20:495-514. [doi: 10.1007/978-3-030-82430-3_21]
44. Urbano ICVP, Guerreiro JPV, Nicolau HMAA. From skeuomorphism to flat design: age-related differences in performance and aesthetic perceptions. Behav Inform Technol 2020;41(3):452-467. [doi: 10.1080/0144929X.2020.1814867]
45. Loorak MH, Perin C, Kamal N, Hill M, Carpendale S. TimeSpan: using visualization to explore temporal multi-dimensional data of stroke patients. IEEE Trans Vis Comput Graph 2016;22(1):409-418 [FREE Full text] [doi: 10.1109/TVCG.2015.2467325] [Medline: 26390482]
46. Sowmia KR, Jayaganeshan T, Khan MA, Madhesh S, Kabilesh S. An artificial intelligence approach to quantifying exercise form for optimal performance and injury prevention. : Springer; 2024 Presented at: Proceedings of Third International Conference on Computing and Communication Networks Lecture Notes in Networks and Systems; 2024; Singapore p. 639-647. [doi: 10.1007/978-981-97-0892-5_50]

XSL•FO
RenderX

47. Ramesh SH, Lemaire ED, Tu A, Cheung K, Baddour N. Automated implementation of the Edinburgh Visual Gait Score (EVGS) using openpose and handheld smartphone video. Sensors (Basel) 2023;23(10):4839 [FREE Full text] [doi: 10.3390/s23104839] [Medline: 37430751]
48. Kotte H, Daiber F, Kravcik M, Duong-Trung N. FitSight: tracking and feedback engine for personalized fitness training. 2024 Presented at: UMAP '24: Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization; 2024 June 22; Cagliari, Italy p. 223-231. [doi: 10.1145/3627043.3659547]
49. Postolache O, Hemanth DJ, Alexandre R, Gupta D, Geman O, Khanna A. Remote monitoring of physical rehabilitation of stroke patients using IoT and virtual reality. IEEE J Select Areas Commun 2021;39(2):562-573. [doi: 10.1109/jsac.2020.3020600]
50. Chatzitofis A, Zarpalas D, Filos D, Triantafyllidis A, Chouvarda I, Maglaveras N, et al. Technological module for unsupervised, personalized cardiac rehabilitation exercising. 2017 Presented at: IEEE 41st Annual Computer Software and Applications Conference (COMPSAC); 2017 July 04-08; Turin, Italy p. 125-130. [doi: 10.1109/COMPSAC.2017.230]
51. Güler RA, Neverova N, Kokkinos I. DensePose: dense human pose estimation in the wild. 2018 Presented at: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition; 2018 June 18-23; Salt Lake City, UT p. 7297-7306. [doi: 10.1109/CVPR.2018.00762]
52. Maji D, Nagori S, Mathew M, Poddar D. YOLO-Pose: enhancing YOLO for multi person pose estimation using object keypoint similarity loss. 2022 Presented at: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2022 June 19-20; New Orleans, LA p. 2636-2645. [doi: 10.1109/CVPRW56347.2022.00297]
53. Cao Z, Hidalgo G, Simon T, Wei S, Sheikh Y. OpenPose: realtime multi-person 2D pose estimation using part affinity fields. IEEE Trans Pattern Anal Mach Intell 2021;43(1):172-186. [doi: 10.1109/TPAMI.2019.2929257] [Medline: 31331883]
54. Rangari T, Kumar S, Roy PP, Dogra DP, Kim B. Video based exercise recognition and correct pose detection. Multimed Tools Appl 2022;81(21):30267-30282. [doi: 10.1007/s11042-022-12299-z]
55. Elsakr C, Bulger DA, Roman S, Kirolos I, Khouzam RN. Barriers physicians face when referring patients to cardiac rehabilitation: a narrative review. Ann Transl Med 2019;7(17):414 [FREE Full text] [doi: 10.21037/atm.2019.07.61] [Medline: 31660313]
56. Ghisi GLM, Polyzotis P, Oh P, Pakosh M, Grace SL. Physician factors affecting cardiac rehabilitation referral and patient enrollment: a systematic review. Clin Cardiol 2013;36(6):323-335 [FREE Full text] [doi: 10.1002/clc.22126] [Medline: 23640785]
57. Pedersen DH, Mansourvar M, Sortsø C, Schmidt T. Predicting dropouts from an electronic health platform for lifestyle interventions: analysis of methods and predictors. J Med Internet Res 2019;21(9):e13617 [FREE Full text] [doi: 10.2196/13617] [Medline: 31486409]
58. Bremer V, Chow PI, Funk B, Thorndike FP, Ritterband LM. Developing a process for the analysis of user journeys and the prediction of dropout in digital health interventions: machine learning approach. J Med Internet Res 2020;22(10):e17738 [FREE Full text] [doi: 10.2196/17738] [Medline: 33112241]
59. Bennemann B, Schwartz B, Giesemann J, Lutz W. Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. Br J Psychiatry 2022;220(4):192-201. [doi: 10.1192/bjp.2022.17] [Medline: 35177132]
60. Lee S, Chung JY. The machine learning-based dropout early warning system for improving the performance of dropout prediction. Appl Sci 2019;9(15):3093. [doi: 10.3390/app9153093]
61. Gross DP, Zhang J, Steenstra I, Barnsley S, Haws C, Amell T, et al. Development of a computer-based clinical decision support tool for selecting appropriate rehabilitation interventions for injured workers. J Occup Rehabil 2013;23(4):597-609. [doi: 10.1007/s10926-013-9430-4] [Medline: 23468410]
62. Lofaro D, Groccia M, Guido R, Conforti D, Caroleo S, Fragomeni G. Machine learning approaches for supporting patient-specific cardiac rehabilitation programs. Comput Cardiol (2010) 2016;43:149-152. [doi: 10.22489/CinC.2016.047-209]
63. Nithya B, Ilango V. Predictive analytics in health care using machine learning tools and techniques. 2017 Presented at: Proceedings of the International Conference on Intelligent Computing and Control Systems, ICICCS; 2017 June 15-16; Madurai, India p. 492-499. [doi: 10.1109/ICCONS.2017.8250771]
64. Valdes G, Simone CB, Chen J, Lin A, Yom SS, Pattison AJ, et al. Clinical decision support of radiotherapy treatment planning: a data-driven machine learning strategy for patient-specific dosimetric decision making. Radiother Oncol 2017;125(3):392-397. [doi: 10.1016/j.radonc.2017.10.014] [Medline: 29162279]
65. Swarnakar R, Yadav SL. Artificial intelligence and machine learning in motor recovery: a rehabilitation medicine perspective. World J Clin Cases 2023;11(29):7258-7260 [FREE Full text] [doi: 10.12998/wjcc.v11.i29.7258] [Medline: 37946764]
66. Fahrmeir L, Kneib T, Lang S, Marx B. Regression models. In: Regression. Berlin, Heidelberg: Springer; 2013:21-72. [doi: 10.1007/978-3-642-34333-9]
67. Yang FJ. An implementation of naive bayes classifier. 2018 Presented at: International Conference on Computational Science and Computational Intelligence (CSCI); 2018 December 12-14; Las Vegas, NV p. 301-306. [doi: 10.1109/CSCI46756.2018.00065]
68. Breiman L. Random forests. Mach Learn 2001;45(1):5-32. [doi: 10.1023/A:1010933404324]

XSL•FO
RenderX

69. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. 2016 Presented at: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016 August 13; San Francisco, CA p. 785-794. [doi: https://doi.org/10.1145/2939672.2939785]
70. Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot 2013;7:21 [FREE Full text] [doi: 10.3389/fnbot.2013.00021] [Medline: 24409142]
71. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. IEEE Intell Syst Their Appl 1998;13(4):18-28. [doi: 10.1109/5254.708428]
72. Duan H, Wang Z, Ji Y, Ma L, Liu F, Chi M, et al. Using goal-directed design to create a mobile health app to improve patient compliance with hypertension self-management: development and deployment. JMIR Mhealth Uhealth 2020;8(2):e14466 [FREE Full text] [doi: 10.2196/14466] [Medline: 32130161]
73. Mozafarinia M, Rajabiyazdi F, Brouillette M, Fellows LK, Mayo NE. Development and usability of a feedback tool, "My Personal Brain Health Dashboard", to improve setting of self-management goals among people living with HIV in Canada. Qual Life Res 2021;30(11):3199-3211. [doi: 10.1007/s11136-020-02555-w] [Medline: 32556825]
74. Ha Y, Park M. Effects of stroke rehabilitation using gait robot-assisted training and person-centered goal setting: a single blinded pilot study. Healthcare (Basel) 2023;11(4):588 [FREE Full text] [doi: 10.3390/healthcare11040588] [Medline: 36833122]
75. Hersh D, Worrall L, Howe T, Sherratt S, Davidson B. SMARTER goal setting in aphasia rehabilitation. Aphasiology 2012;26(2):220-233. [doi: 10.1080/02687038.2011.640392]
76. Chaix B, Guillemassé A, Nectoux P, Delamon G, Brouard B. Vik: a chatbot to support patients with chronic diseases. Health N Hav 2020;12(07):804-810. [doi: 10.4236/health.2020.127058]
77. Zand A, Sharma A, Stokes Z, Reynolds C, Montilla A, Sauk J, et al. An exploration into the use of a chatbot for patients with inflammatory bowel diseases: retrospective cohort study. J Med Internet Res 2020;22(5):e15589 [FREE Full text] [doi: 10.2196/15589] [Medline: 32452808]
78. Bibault J, Chaix B, Guillemassé A, Cousin S, Escande A, Perrin M, et al. A chatbot versus physicians to provide information for patients with breast cancer: blind, randomized controlled noninferiority trial. J Med Internet Res 2019;21(11):e15787 [FREE Full text] [doi: 10.2196/15787] [Medline: 31774408]
79. Khan V, Meenai TA. Pretrained natural language processing model for intent recognition (BERT-IR). Hum Centric Intell Syst 2021;1(3):66-74. [doi: 10.2991/hcis.k.211109.001]
80. Finkel J, Grenager T, Manning C. Incorporating non-local information into information extraction systems by Gibbs sampling. 2005 Presented at: ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics; June 25, 2005; Ann Arbor, MI p. 363-370. [doi: 10.3115/1219840.1219885]
81. Akbik A, Blythe D, Vollgraf R. Contextual string embeddings for sequence labeling. 2018 Presented at: Proceedings of the 27th International Conference on Computational Linguistics; 2018 August 20-26; Santa Fe, NM p. 1638-1649 URL: https://aclanthology.org/C18-1139/
82. Glazkova A, Glazkov M, Trifonov T. g2tmn at Constraint@AAAI2021xploiting CT-BERT and Ensembling Learning for COVID-19 Fake News Detection. Commun Comput Inf Sci 2021:116-127. [doi: 10.1007/978-3-030-73696-5_12]
83. Devlin J, Chang M, Lee K, Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. 2018 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); June 01, 2019; Minneapolis, MN p. 4171-4186. [doi: 10.1038/s41569-021-00611-7]
84. Doran G. There's a S.M.A.R.T. way to write managements's goals and objectives. Manage Rev 1981;70(11):35-36 [FREE Full text]
85. Anandan P, Kokila S, Elango S, Gopinath P, Sudarsan P. Artificial intelligence based chat bot for patient health care. 2022 Presented at: International Conference on Computer Communication and Informatics (ICCCI); 2022 January 25-27; Coimbatore, India. [doi: 10.1109/ICCCI54379.2022.9740912]

Abbreviations

COREQ: COnsolidated criteria for REporting Qualitative research
CR: cardiac rehabilitation
EMR: electronic medical record
LLM: large language model
ML: machine learning
NLP: natural language processing
SMART: Specific, Measurable, Achievable, Relevant, Time-bound
VCR: virtual cardiac rehabilitation

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 05.09.24; peer-reviewed by M Mohammadnezhad, J Walsh; comments to author 04.12.24; revised version received 15.01.25; accepted 21.01.25; published 21.02.25.

Please cite as:
Ramesh SH, Jull D, Fournier H, Rajabiyazdi F
Exploring Barriers to Patients' Progression in the Cardiac Rehabilitation Journey From Health Care Providers' Perspectives: Qualitative Study
Interact J Med Res 2025;14:e66164
URL: https://www.i-jmr.org/2025/1/e66164
doi:10.2196/66164
PMID:

©Shri Harini Ramesh, Darwin Jull, Hélène Fournier, Fateme Rajabiyazdi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 21.02.2025. 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 Dynamics of Inhibitors and Facilitators of Exercise Behavior Within the Transtheoretical Model: Nationwide Cross-Sectional Study Using Text Mining Analysis

Kosuke Sakai[1], MD, PhD; Kota Fukai[1], MD, PhD; Yuko Furuya[1], MD, PhD; Shoko Nakazawa[1], MD, PhD; Kei Sano[2], MD, PhD; Masayuki Tatemichi[1], MD, PhD

[1]Department of Preventive Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa, Japan

[2]Department of Ophthalmology, Jikei University School of Medicine, Tokyo, Japan

Corresponding Author:
Kota Fukai, MD, PhD
Department of Preventive Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa, Japan

Abstract

Background: The transtheoretical model (TTM) explains behavior changes through sequential stages influenced by the balance of perceived benefits and barriers. Although previous studies have identified the inhibitors and facilitators of exercise behavior, only a few have elucidated how these factors vary across the stages of behavior change.

Objective: This study aimed to identify the inhibitors and facilitators of each stage of behavior change using text mining.

Methods: A nationwide cross-sectional study was conducted using an internet-based questionnaire with a panel of approximately 2 million members. From this panel, 93,460 individuals were randomly selected and invited to participate via email and app push notifications. A total of 1500 valid responses were included in the analysis through stratified sampling based on sex, age group, and geographic region. The participants self-assessed their stages of change. Two open-ended questions captured the perceptions of inhibitors and facilitators of exercise behavior. Text responses were analyzed in a 4-step process: morphological analysis to extract frequently used words, correspondence analysis to visualize relationships between frequently used words and the 5 change stages, conceptual categorization with coding rules, and creation of heat maps to illustrate stage-specific categories in inhibitors and facilitators.

Results: Out of 1500 respondents, 754 (50.3%) were males and 355 (23.7%) individuals were in the 50 - 59 age group. Stage percentages were precontemplation 24.3% (365/1500), contemplation 23.5% (352/1500), preparation 21.3% (320/1500), action 5% (75/1500), and maintenance 25.9% (388/1500). The inhibitors and facilitators were described using 9893 words and 8372 words, respectively. Inhibitors clustered into 7 categories; most frequent were time (408/1500, 27.2%), motivation (253/1500, 16.9%), health (189/1500, 12.6%), and working (158/1500, 10.5%). Facilitators formed 8 categories; most frequent were subjectivity (155/1500, 10.3%), relationship (93/1500, 6.2%), opportunity (84/1500, 5.6%), reward (78/1500, 5.2%), and record (78/1500, 5.2%). Stage-specific patterns emerged: inhibitors shifted from motivation and health (precontemplation) to family, time, and working (contemplation and preparation), opportunity (action), and weather and health (maintenance). Facilitators of reward, health, and record rose progressively from precontemplation to maintenance.

Conclusions: This study enhances our understanding of the dynamic mechanisms underlying exercise behavior change by identifying how specific inhibitors and facilitators vary across behavioral stages. The findings underscore the need to tailor interventions based on individuals' readiness to change, rather than relying on one-size-fits-all strategies. For both practitioners and policymakers, incorporating behavioral stage frameworks into assessments and interventions, such as those conducted during health checkups or workplace programs, may improve the effectiveness and sustainability of physical activity promotion.

(Interact J Med Res 2025;14:e77400) doi:10.2196/77400

KEYWORDS

transtheoretical model; text mining; exercise; inhibitor; facilitator

Introduction

Exercise habits can improve mental and cardiovascular health [1,2]. A key health policy guideline for Japan, "Health Japan 21," placed a high priority on exercise habits, with the goal of increasing the percentage of people with exercise habits from 28.7% in 2019 to 40% by 2032 [3]. In this guideline, "exercise habits" are defined as engaging in exercise for at least 30 minutes per session, at least twice a week, and continuing for at least 1 year. Despite this policy target, the exercise habit rate has decreased to 27.3% in 2023 due to the coronavirus disease

XSL•FO
RenderX

2019 epidemic [4]. The focus is on providing effective interventions to promote exercise habits.

The behavior change process is explained using the transtheoretical model (TTM) [5,6]. The model categorizes the process of behavior change into 5 stages—precontemplation, contemplation, preparation, action, and maintenance—assigns individuals to one of these stages, and conceptualizes each stage as dynamically evolving over time. Various factors are associated with the acquisition of exercise habits, including individual characteristics, societal norms, culture, and the environment [7]. In the TTM, the psychologically influential factors are categorized into cognitive process of change, intrinsic mental shifts, and behavioral process of change, which are influenced by environmental and external support [8]. In addition, a person's stage can change depending on the balance between perceived benefits and barriers to exercise. Interventions designed using the TTM have been shown to be highly effective in changing behaviors [9]. Unlike other models that focus solely on intention or beliefs, such as the health belief model or the theory of planned behavior, the TTM is uniquely suited to capture temporal changes and stage-specific factors. This makes it particularly appropriate for identifying which facilitators and inhibitors are most relevant at each stage of exercise behavior. Furthermore, in Japan, the TTM has been applied in public health efforts such as the implementation of "tokutei-kenko-shinsa" (specific health checkups), a nationwide preventive health screening program targeting metabolic syndrome. Thus, the TTM has both theoretical and practical relevance for understanding and promoting exercise behavior in the Japanese population [10].

Factors influencing behavior change can be divided into inhibitors and facilitators [11]. Inhibitors include diseases, psychological barriers, time constraints, and poor access to exercise environments. Facilitators include self-efficacy, positive feelings toward exercise, good health, and social support [11]. The extent to which these factors influence exercise behavior varies depending on the individual stages of behavior change [12]. A previous study has found that physical, mental, and environmental barriers have a significant influence on the precontemplation and contemplation stages, perceived behavioral control is significant in the contemplation stage, and self-efficacy and social support play a significant role in the action and maintenance stages [13]. The relationship between these factors and the stages of behavior change is complex [14].

Recently, text mining has been recognized as an effective method for elucidating increasingly complex phenomena [15,16]. Previous studies aiming to identify inhibitors and facilitators of exercise habit have collected qualitative data through semistructured interviews and focus group discussions, which were then analyzed and interpreted by researchers to identify and categorize key concepts [11,13,17,18]. Text mining enables the visualization and quantitative evaluation of relationships between words in free-text responses and respondents' demographic characteristics. Text mining has been used in various public health studies, leveraging its high reproducibility and capacity to process large volumes of data [19,20]. Studies using text mining to analyze the influential factors in the exercise habits have been limited to college students and workers in specific occupations [21,22]. To the best of our knowledge, no previous studies have used this method to analyze the relationship between behavior change stages and both inhibitors and facilitators. Text mining may yield valuable insights into the inhibitors and facilitators, as well as elucidate the complex relationships between these factors and their respective stages.

This study aimed to identify the inhibitors and facilitators of each stage of behavior change using text mining. Unraveling the complexity of the factors influencing behavior change will deepen our understanding of exercise habit formation and support the development of individualized interventions.

Methods

Study Design and Participants

This was a nationwide cross-sectional study in Japan using an internet questionnaire survey, which was administered in November 2024. The survey was conducted by Rakuten Insight Inc, a web-based research firm in Tokyo, Japan. It was one of the largest and most widely used web-based research providers in the country, with a panel of approximately 2 million members. The company has also been used in previous peer-reviewed public health research [23,24]. This study was reported in accordance with the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) guidelines for cross-sectional observational studies (Checklist 1) and the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) checklist for Internet-based surveys (Checklist 2) [25,26].

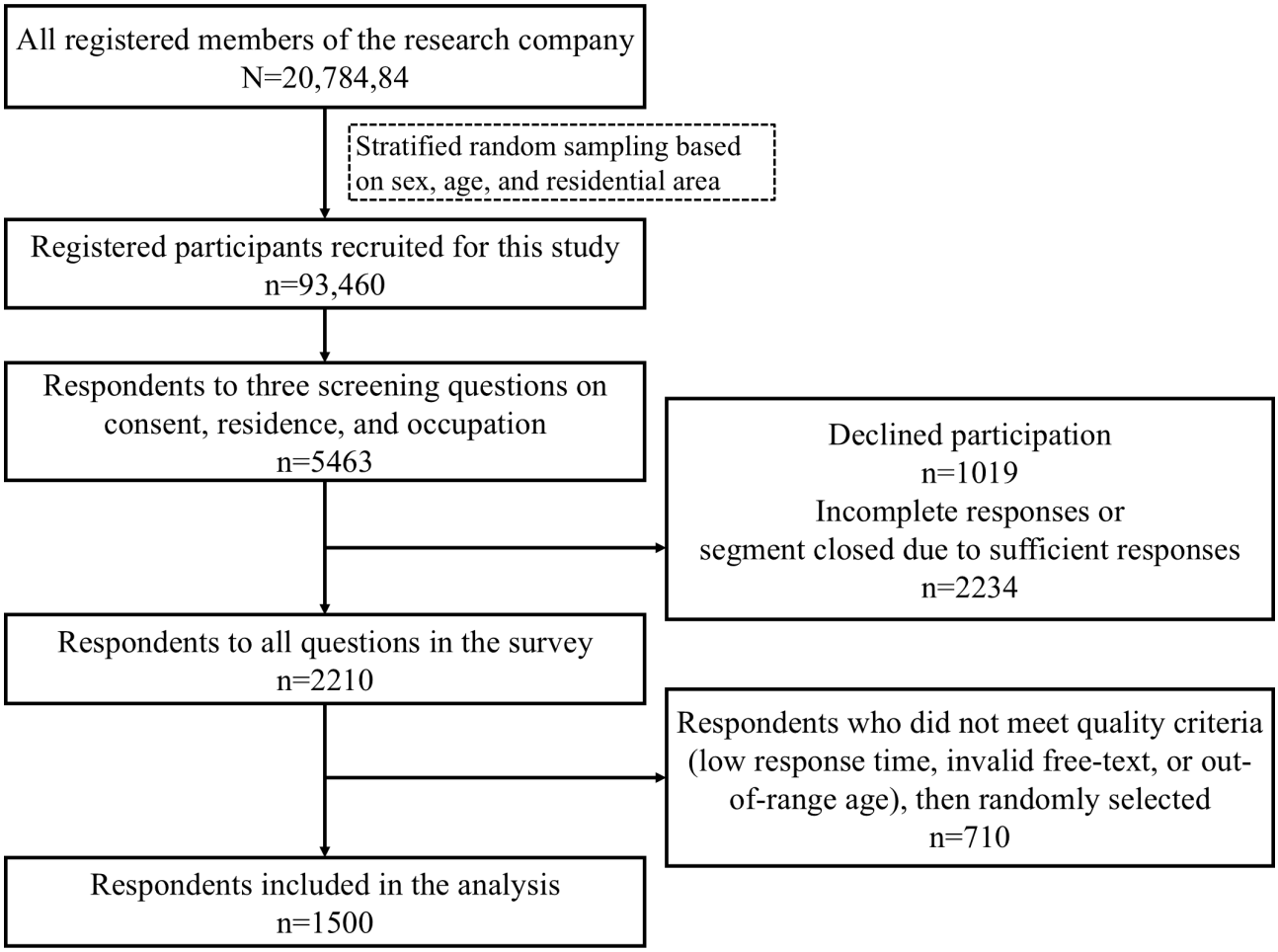
The participants were recruited from the company's registry. The inclusion criterion was aged 20 - 69 years living in Japan. To ensure a representative sample, stratified sampling was used across 200 segments based on sex (male and female), 5-year strata age groups, and 10 geographical regions (Hokkaido, Tohoku, North Kanto, South Kanto, Hokuriku, Tokai, Kinki, Chugoku, Shikoku, and Kyushu) in accordance with the distribution of population estimates from the 2020 census [27]. Surveys were conducted until 1500 valid responses were obtained, with target sample quotas established for each segment. The sample size was determined pragmatically as the maximum number of responses that could be obtained within the available budget. The questionnaire consisted of one question per page, totaling 21 pages. Respondents were able to review and change their answers before final submission.

Figure 1 illustrates the process of selecting participants for the final analysis. From a pool of 2,078,484 registered members of the survey panel, 93,460 participants were invited to take part in the study via email and push notifications through the company's app. These participants were selected using stratified random sampling based on sex, age, and residential area. Of these, 5463 individuals responded to 3 initial screening questions regarding their consent to participate, place of residence, and occupation. A total of 3253 individuals were excluded at this stage due to declining participation, providing incomplete responses, or belonging to quota-saturated demographic segments. Among the 2210 individuals who responded to the

XSL•FO
RenderX

survey, 1500 were randomly selected for analysis using the following procedure. First, a validity flag was assigned to samples that met predefined criteria. Before flagging, respondents were excluded if they (1) had response times below a threshold (defined as less than the median response time multiplied by 0.3 for PC users or 0.2 for smartphone users), (2) provided clearly inappropriate answers in free-text fields, or (3) reported an age outside the target range (18 - 79 y). Next, a random number was assigned to each valid sample using the RAND function in Excel, which generates a random number between 0 and 1. The samples were then sorted in ascending order based on the random number, and the top 1500 were selected for inclusion in the final dataset. Cookies and IP addresses were not used in this survey. This was a closed survey conducted through a commercial Internet survey company. Respondents were required to log in with their unique panel ID before accessing the questionnaire. Once a panelist had completed the survey, the questionnaire was no longer displayed upon subsequent logins, thereby preventing duplicate entries.

Figure 1. Flowchart of participant selection for analysis.



Assessment of Behavior Change Stage in the Transtheoretical Model

Behavior change stage was measured using the Japanese version of the Stages of Exercise Behavior Change Scale, which has been validated in Japanese populations and shown to have acceptable reliability and construct validity [28]. This required respondents to select the appropriate item from the following 5 options: precontemplation, "I currently do not exercise and do not intend to exercise for the next 6 months;" contemplation, "I currently do not exercise, but I intend to exercise within the next 6 months;" preparation, "I currently exercise but not regularly;" action, "I currently exercise regularly, but I have only begun to do so within the past 6 months;" and maintenance, "I currently exercise regularly and have been doing so for >6 months." Regular exercise was defined as exercise lasting >20 - 30 min per set, 2 - 3 times per week.

Assessment of Inhibitors and Facilitators for Exercise

We collected free responses regarding the inhibitors and facilitators of exercise using 2 open-ended questions. To elicit open-ended responses, participants were presented with the following prompts: (1) For inhibitors: "What kinds of obstacles do you face in continuing your exercise habits? Please describe them in as much detail as possible." (2) For facilitators: "What kinds of things have helped or supported you in continuing your exercise habits? Please describe them in as much detail as possible."

Participants were instructed to list as many ideas as they could think of and were encouraged to be specific. These questions were mandatory, and respondents could not proceed to the next section without entering a response.

XSL•FO
RenderX

Assessment of Respondents' Characteristics

The demographic information of the respondents, including sex (male and female), age (20 - 69 y) at the date of response as identified by date of birth, marital status (married, never married, and divorced or deceased), education (junior or high school, vocational school or junior college, university, graduate school, currently enrolled, and other), and household income (<2.0, 2.0 - 3.9, 4.0 - 5.9, 6.0 - 7.9, 8.0 - 9.9, 10.0 - 13. 9, and ≥14.0 million Japanese yen; US $1 = ¥150, based on the average exchange rate in November 2024), was collected.

Statistical Analyses

The text mining analyses were performed in 4 steps. Statistical analysis was performed using KH-coder (version 3.02c official package, SCREEN Advanced System Solutions Co, Ltd) [29]. KH Coder was initially developed as an open-source software for academic research and gained wide recognition due to its accessibility and functionality. Over time, its analytical methods were validated through numerous peer-reviewed publications across fields such as public health and social sciences [16,19,20,30]. Its core methods—such as morphological analysis, co-occurrence networks, and correspondence analysis—have been repeatedly used to analyze publicly available text data, thereby ensuring transparency and reproducibility. The consistency of results reported across independent studies provides evidence supporting the methodological validity and reliability of this software. We analyzed the descriptions in Japanese and translated the words into English.

Morphological Analysis

In the first step, we performed a morphological analysis to determine frequently used words, their parts of speech, and the number of occurrences using the text-mining dictionary ChaSen Morphological Analyzer, which was the default engine in KH Coder and has been widely selected for its compatibility and proven utility in academic research settings [16,20].

Correspondence Analysis

In the second step, we performed a correspondence analysis to visualize the relationship between the top 40 most frequent words and the 5 stages of change, stratified by inhibitors and facilitators. Correspondence analysis was classified as a multivariate analysis and selected when the variables were categorical data. This method aids in visually understanding the relationships among categorical data, which can be represented using a 2D scatterplot [31]. Categories located farther from the origin and closer to the direction of self-rated health levels were considered more specific. Correspondence analysis is often used in marketing to visualize the associations between multiple categorical variables and consumer characteristics. In the field of public health, we have applied this method in our previous research to identify key personnel responsible for employee health management in the workplace and to determine appropriate marketing strategies through various media channels [30].

Conceptualization and Coding Rules

In the third step, we conceptualized the descriptions into several categories from the frequent word lists of inhibitors and facilitators. The coding rules were developed through an iterative process by the first author (KS) who has published multiple peer-reviewed studies using text mining, in collaboration with KF. To ensure the contextual validity of the coding rules, all free-text responses were read in full before conducting morphological analysis. After extracting the frequently used words through morphological analysis, 2 researchers independently reviewed all responses containing those words to understand the specific contexts in which each word was used. Similar words were grouped together, and each group was assigned a conceptual label that reflected the underlying theme. After drafting the initial coding rules, the researchers revisited all responses in which those words appeared to confirm the appropriateness of the assigned categories. Words that were used in multiple or ambiguous contexts were excluded from the coding rules to preserve specificity. After a 1-month interval, the researchers repeated the process to refine and finalize the coding scheme, thereby improving consistency and minimizing potential bias. These processes were carried out through discussions between the 2 researchers. In cases where disagreements arose during the categorization process, the supervising author (MT) was consulted to provide a third opinion and reach consensus. Coding rules were established for each category based on words that appeared at least 10 times. We excluded stop words, such as conjunctions and prepositions and words used in a variety of contexts, such as Nothing, Not, and Exercise. We calculated the number of responses for each category based on coding rules.

Heat Maps

In the fourth step, we created heat maps to visualize how each category varied across the stages of change. The heat maps visualized the variation in the mentions of each category using Pearson's residual scales. Percentages were calculated using the number of responses in each stage as the denominator and the number of responses mentioning each category as the numerator. Pearson residuals indicate the degree to which observed values deviate from expected values in the analysis of mentions in a particular category. The heat map displays the color intensity when mentions of a particular category occur more or less frequently than expected.

$$\text{Pearson residual} = (O - E)/\sqrt{E}$$

O: observational number of responses mentioning the category.

E: expected number of responses mentioning the category.

The process of developing coding rules and visualizing conceptual categories through heat maps has been effectively applied in our previous studies, such as an investigation into the supportive behaviors of supervisors with high work engagement and an analysis of trends in occupational health research over the past three decades [16,19]. These applications have demonstrated the utility and validity of the approach.

XSL•FO
RenderX

Sensitivity Analysis

To assess the robustness of the findings, we conducted sensitivity analyses stratified by sex (males and females) and education (less than university and university or more). The same coding rules and analytical procedures described above were applied separately to each group to examine whether the stage-specific distribution patterns of inhibitors and facilitators were consistent.

Ethical Considerations

This study was approved by the Research Ethics Committee of the Tokai University School of Medicine (24R109). Informed consent was obtained from all the respondents before their participation in the questionnaire survey. To incentivize participation, respondents who passed the screening questions received 2 points, and those who completed the full survey received from 15 to 17 points. These points were equivalent to 1 yen per point and could be exchanged for merchandise on an affiliated online platform. All responses were anonymized by the survey company before being provided to the researchers, and no personally identifiable information was collected.

Results

Characteristics of Respondents

Respondents' characteristics are listed in Table 1. Of 1500 respondents, 754 (50.3%) were males. In addition, 23.6% (178/754) males and 23.7% (177/746) females were in their 50 s. A total of 50.8% (383/754) males and 56.7% (423/746) females were married. Household incomes between 4 million and 5.99 million yen (US $1 = ¥150) were reported by 30.1% (227/754) men and 23.5% (175/746) women. In terms of the stage of change, 29.6% (223/754) men were in maintenance, 23.9% (180/754) in preparation, and 22.5% (170/754) in precontemplation; 28.6% (213/746) women were in contemplation, 26.1% (195/746) in precontemplation, and 22.1% (165/746) in maintenance. Because all questionnaire items were mandatory, there were no missing values among the analyzed participants.

XSL•FO
RenderX

Table . Characteristics of respondents.

	Males, n (%)	Females, n (%)
Total	754 (100)	746 (100)
Age group (years)		
20 - 29	129 (17.1)	122 (16.4)
30 - 39	135 (17.9)	131 (17.6)
40 - 49	170 (22.5)	165 (22.1)
50 - 59	178 (23.6)	177 (23.7)
60 - 69	142 (18.8)	151 (20.2)
Marital history		
Married	383 (50.8)	423 (56.7)
Divorce or bereavement	42 (5.6)	85 (11.4)
Unmarried	329 (43.6)	238 (31.9)
Education		
Junior, high school	206 (27.3)	208 (27.9)
Vocational school, junior college	105 (13.9)	268 (35.9)
University	364 (48.3)	249 (33.4)
Graduate school	57 (7.6)	9 (1.2)
Currently enrolled	18 (2.4)	11 (1.5)
Other	4 (0.5)	1 (0.1)
Household income, million yen[a]		
<2.0	75 (9.9)	109 (14.6)
2.0 to 3.9	135 (17.9)	181 (24.3)
4.0 to 5.9	227 (30.1)	175 (23.5)
6.0 to 7.9	128 (17)	116 (15.5)
8.0 to 9.9	82 (10.9)	78 (10.5)
10.0 to 13.9	69 (9.2)	54 (7.2)
≥14.0	38 (5)	33 (4.4)
Stage of change		
Precontemplation	170 (22.5)	195 (26.1)
Contemplation	139 (18.4)	213 (28.6)
Preparation	180 (23.9)	140 (18.8)
Action	42 (5.6)	33 (4.4)
Maintenance	223 (29.6)	165 (22.1)

[a]150 Japanese yen = US $1.

Morphological Analysis

In the first step, we detected inhibitors using 9893 extracted words and 1157 unique words and facilitators using 8372 extracted words and 1241 unique words. Frequent words described for inhibitors, those appearing at least 10 times, are shown in Multimedia Appendix 1, and frequent words described for facilitators are shown in Multimedia Appendix 2. The inhibitors included time (N=387), job (N=154), exhausted (N=58), eagerness (n=49), troublesome (n=46), and busy (n=44). The facilitators included nothing (n=542), gymnasium (n=56), time (n=52), together (n=48), understand (n=44), application (n=38), and goal (n=32).

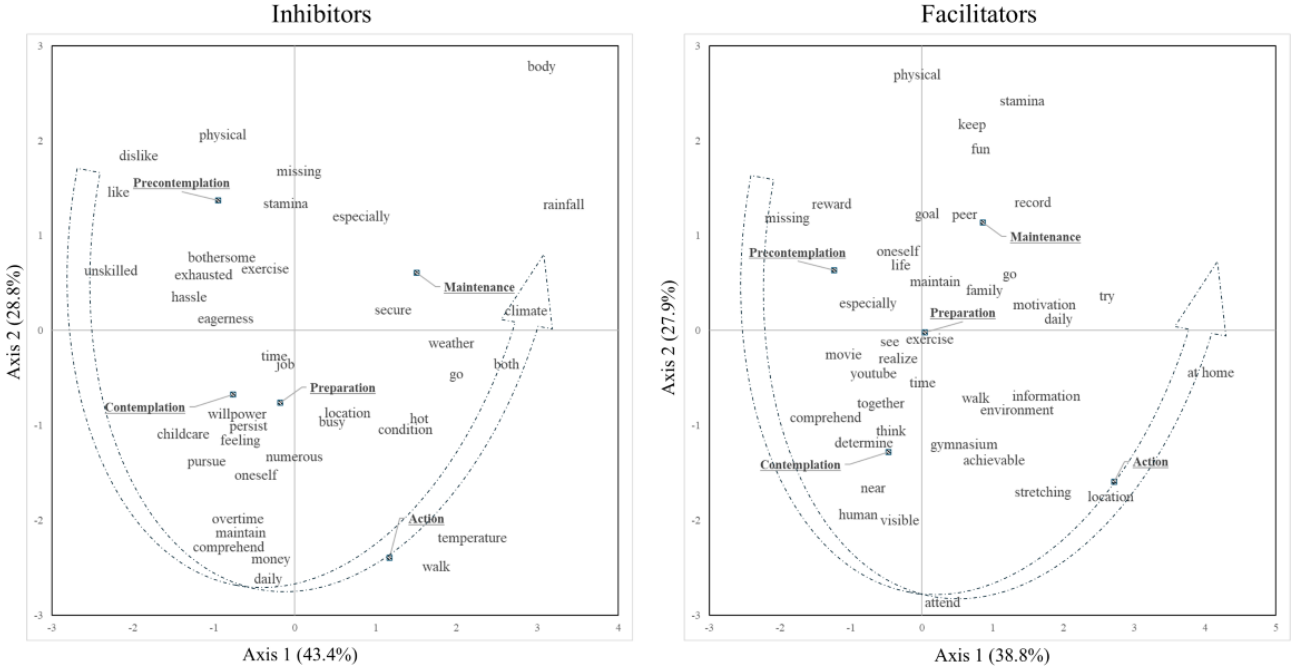
Correspondence Analyses

In the second step, Figure 2 presents the results of the correspondence analyses to visualize the relationship between the stages of change and the 40 specific words in the inhibitors and facilitators. The stages of change were arranged counterclockwise for both the inhibitors and facilitators. A spectrum was drawn from the appearance of the word characteristics at each stage. In the inhibitors, dislike, like,

XSL•FO
RenderX

bothersome, and unskilled appeared in the precontemplation period; childcare, willpower, overtime, and money appeared in the contemplation and preparation; walk and temperature appeared in the action period; and body, rainfall, and climate appeared in the maintenance period. In the facilitators, missing and reward emerged in precontemplation; near, human, and visible emerged in contemplation; location and stretching emerged in action; and stamina, fun, record, and peer emerged in maintenance.

Figure 2. This figure presents the results of the correspondence analysis, illustrating the relationship between the stages of behavioral change and frequently mentioned words related to inhibitors (left panel) and facilitators (right panel) of exercise habits.



Conceptualization and Coding Rules

In the third step, the inhibitors and facilitators were structured into several categories using coding rules, as shown in Table 2. The inhibitors were divided into seven categories (motivation, health, weather, family, time, opportunity, and working). In response to the inhibitors, 408 (27.2%) respondents referred to the category of time, including words, such as time, busy, and capacity. Similarly, 253 (16.9%) respondents mentioned motivation, 189 (12.6%) mentioned health, and 158 (10.5%) mentioned work. A total of 572 (38.1%) participants did not use any of the words included in the coding rules. After reading all the responses in the category of no coding, we observed that 255 respondents answered, “I do not know” or “nothing in particular.” The facilitators comprised 8 categories (reward, subjectivity, opportunity, relationship, health, record, digital device, and time). Among the facilitators’ responses, 155 (10.3%) referred to the category of subjectivity, which included words such as understand, achievable, comprehend, and determine. Similarly, 93 (6%) and 84 (6%) participants mentioned the relationship and opportunity categories, respectively. There were 956 (63.7%) responses not referring to any categories (no coding), and 617 of these responses had no meaning, such as “I do not know” or “nothing in particular.”

XSL•FO
RenderX

Table . Conceptualization, coding rules, and reference conditions on inhibitors and facilitators.

Factors and categories		Words in coding rules	Reference counts and rate, n (%)
Inhibitors			
	Motivation	eagerness, troublesome, hassle, motivation, feeling, bothersome, willpower, like, dislike, unskilled	253 (16.9)
	Health	exhausted, stamina, condition, physical, knee, fatigue, backache, body	189 (12.6)
	Weather	weather, climate, rainfall, hot, chilly, temperature	98 (6.5)
	Family	chores, parenting, caregiving, childcare	71 (4.7)
	Time	time, busy, capacity	408 (27.2)
	Opportunity	gymnasium, location, money	48 (3.2)
	Working	job, overtime	158 (10.5)
	No coding[a]		572 (38.1)
Facilitators			
	Reward	point, motivation, fun, effect, reward, capable	78 (5.2)
	Subjectivity	understand, achievable, comprehend, determine	155 (10.3)
	Opportunity	gymnasium, location, environment, institution	84 (5.6)
	Relationship	together, peer, human, family	93 (6.2)
	Health	health, stamina, body, physical	68 (4.5)
	Record	goal, weight, record, strides	78 (5.2)
	Digital device	application, YouTube, movie	70 (4.7)
	Time	time, job	68 (4.5)
	No coding[a]		956 (63.7)

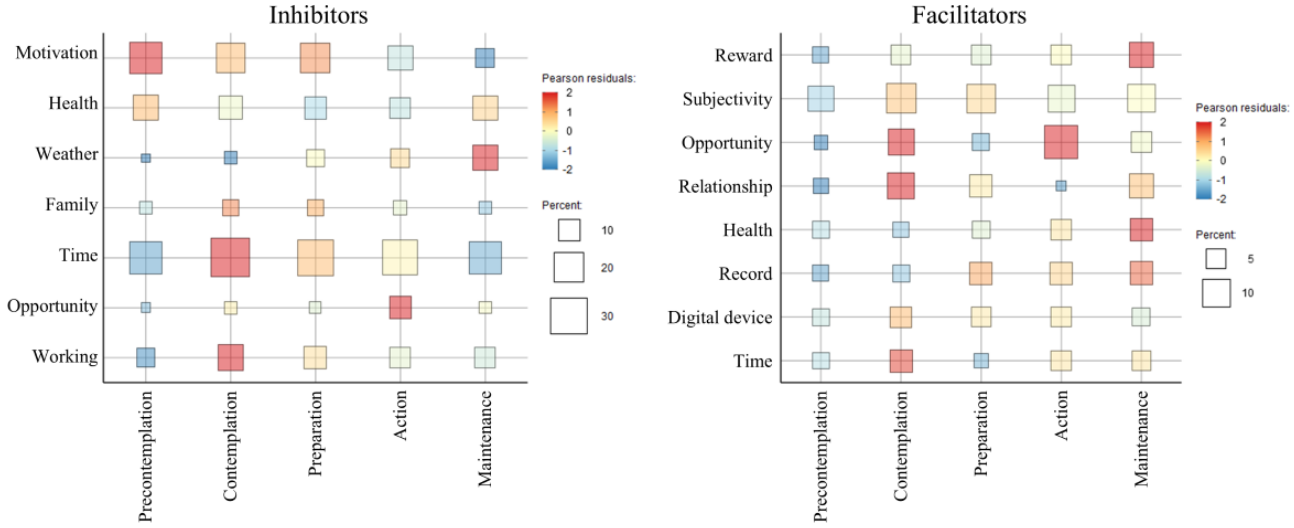
[a]No coding: responses that did not use any of the words listed in the coding rules above.

Relationship Between the Stages of Change and the Categories of Inhibitors and Facilitators

In the fourth step, the heat map in Figure 3 shows the association between the stages of change and the categories of inhibitors and facilitators. The color indicates the bias of the distribution in each category in the horizontal direction, and the size of the square indicates the percentage of mentions at each stage in the vertical direction. In terms of inhibitors, motivation decreased and weather increased as the behavior change stage progressed. The categories of family, time, and work were mentioned in contemplation and preparation, whereas opportunity was mentioned in action. Among the facilitators, reward, health, and record increased as the behavior change stage progressed. Opportunity was mentioned in contemplation and action, whereas relationship and time were mentioned in contemplation. The numbers and percentages of mentions of the categories by stage are detailed in Multimedia Appendix 3.

XSL•FO
RenderX

Figure 3. This heat map illustrates the relationship between the stages of behavioral change and the categories of inhibitors (left) and facilitators (right).



	Precontemplation	Contemplation	Preparation	Action	Maintenance
Inhibitors	Motivation and health	Time, working, and family	Time, working, and family	Opportunity	Weather and health
Facilitators	–	Opportunity, relationship, and time	Record	Opportunity	Reward, health, and record

The results of the sensitivity analysis stratified by sex and education are presented in Multimedia Appendix 4 (inhibitors by sex), Multimedia Appendix 5 (facilitators by sex), Multimedia Appendix 6 (inhibitors by education), and Multimedia Appendix 7 (facilitators by education). For inhibitors, similar trends in category–stage associations were observed, even when the data were stratified by sex and educational attainment. For facilitators, the frequency of mentions across all categories remained consistently low in the precontemplation stage across all sensitivity analyses. However, beyond the precontemplation stage, the associations between facilitator categories and stages of change varied by sex and education, and no consistent patterns were observed.

Discussion

Principal Findings

In this study, a text-mining analysis of open-ended responses about the inhibitors and facilitators of exercise behavior revealed a complex relationship between the stages of change and these factors. Inhibitors were structured into 7 categories (motivation, health, weather, family, time, opportunity, and working) and facilitators into 8 categories (reward, subjectivity, opportunity, relationship, health, record, digital device, and time). The inhibitors included motivation and health in the precontemplation; time, working, and family in the contemplation and preparation; opportunity in the action; and weather and health in the maintenance. As precontemplation progressed to action, reward, health, and record emerged as facilitators. The findings of this study will contribute to a better understanding of the complex mechanisms underlying exercise habit formation and will inform the development of personalized interventions.

The following inhibitors changed as the behavior change stage progressed: motivation, health, flexibility (time, working, and family), opportunity, and weather. Overcoming these inhibitors makes it possible to acquire exercise habits. Similar inhibitors have been reported previously. A systematic review of the determinants of physical activity among adults in the United Kingdom identified lack of time, health problems, and language barriers as inhibitors [32]. A study of individuals with musculoskeletal disorders identified health conditions, time constraints, psychological barriers, and limited access to opportunities, including weather, as inhibitors [11]. To support the acquisition of exercise habits, these inhibitors need to be overcome. In a systematic review and meta-analysis, interventions involving motivational interviewing have been reported to be effective in promoting physical activity and exercise and reducing sedentary time [33]. Motivational interviewing of patients with cancer has reported effects on the acquisition of exercise habits over a period of ≥6 months [34]. Continuous sessions with a physiotherapist and motivational interventions are effective in improving exercise adherence in older adults with chronic low back pain to improve exercise adherence [35]. However, interventions using digital devices that have emerged in recent years have not been proven effective in promoting exercise in patients with musculoskeletal pain [36]. These results indicate the need for ongoing support from specialists. Time management and opportunities such as work–life balance, access to gyms, and financial resources can contribute to the development of exercise habits only when individuals are in good health and sufficiently motivated to exercise.

In the correspondence analyses, the facilitators showed no more fluid relationships with the behavior change stage than the inhibitors. In addition, the results of sensitivity analysis stratified by sex and education for the relationship between facilitators and stages were inconsistent. One possible reason for this is that facilitators vary according to the individuals' backgrounds. In a qualitative study of individuals with chronic back pain, professional supervision and group practice served as facilitators, because pain was the primary inhibitor [18]. For

XSL•FO
RenderX

those with depressive symptoms, emotions, beliefs about capability, and intentions were barriers; therefore, improving the emotional domain was a facilitator [37]. The facilitators depended on the inhibitors possessed by individuals.

As the behavior change stage progressed, reward, health, and record emerged as facilitators. Recording this process is expected to facilitate the observation of changes in the acquisition of exercise habits. As individuals build stamina, notice physical changes, and experience positive emotional rewards, they may progress to the maintenance stage. Setting goals and tracking weight and step counts enhance intervention effectiveness [38]. A cycle was observed in which self-efficacy increases as individuals progress through the behavior change stages, further boosting their motivation to exercise [38]. Tracking progress may act as a trigger for this cycle.

Facilitators were more frequently described in relation to behavioral processes that involve specific actions to modify and maintain behavior rather than experiential processes that focus on changing thoughts, emotions, and awareness of behavior. As previously mentioned, the process of change in the TTM can be categorized into experiential and behavioral aspects, with the latter having a stronger association with physical activity [39]. Among the facilitators in this study, behavioral aspects can be classified into the following categories: reward, opportunity, relationship, record, digital device, and time. Without efforts to develop exercise habits, facilitators are likely to remain unrecognizable. All categories in the heat map of facilitators had few mentions of precontemplation. In contrast, mentions of facilitators increased during the behavior change stage following the contemplation stage, suggesting that respondents can write about the facilitators based on their own experiences. An interview study examining the relationship between behavior change stages and influencing factors reported more statements about barriers and fewer about the benefits of exercise during precontemplation [40]. A telephone study of middle-aged Australian women reported that people in the precontemplation period lacked knowledge and family support [41]. Based on these results, interventions during the precontemplation stage should focus on providing knowledge and increasing awareness.

Applications and digital devices were mentioned as facilitating factors. Digital health interventions are becoming increasingly prevalent in exercise routines and positively associated with increased physical activity [42]. Mechanisms are being established to acquire user data through digital devices and to further encourage behavior change [38]. Text mining is actively used in this context. For example, text-mining analysis of messages from sexual and reproductive health information services in Kenya revealed distinct patterns of interest based on sex and age [43]. This study provides valuable insights into tailoring information dissemination based on the attributes of the target audience. Further studies are required to clarify the mechanisms of behavior change through digital devices.

Strengths and Limitations

This study has several strengths. First, it visualizes fluid variations in the inhibitors and facilitators during the process of acquiring exercise habits. Inhibitors that need to be overcome step-by-step to acquire exercise habits were identified. Second, this study offers a new perspective for evaluating effective exercise interventions. The findings of this study can be used to develop intervention strategies based on inhibitors and facilitators. A meta-analysis of intervention studies on exercise habit formation found that interventions tailored to an individual's stage of behavior change were not necessarily more effective than those that were not [43]. Furthermore, interventions focusing on self-efficacy are more effective than those based solely on behavior change stages. Interventions should consider the stage of change and factors, such as decisional balance, self-efficacy, and the process of change [44]. Third, text mining was used to clarify the behavior change process. A text-mining study analyzing free-text descriptions of exercise motivation among college students found that the high-motivation group referred to enjoyment, sports, and participation, whereas the low-motivation group mentioned dislike, rivals, and partners [21]. A previous study on coal miners applied latent Dirichlet allocation with a topic model to identify key factors, including physical and psychosocial environments, individual characteristics, and health beliefs [22]. Unlike previous studies that either focused on comparing subgroups (high vs low motivation) or identifying general themes using topic modeling, our study integrated multiple analytical approaches to comprehensively examine the dynamics of behavior change. Specifically, we applied correspondence analysis to visualize fluid transitions in frequently mentioned words across the stages of change, enabling the identification of stage-specific language patterns. Furthermore, we interpreted the extracted terms and inductively developed conceptual categories, which were then linked with participants' behavioral stages through heat maps. This multifaceted approach allowed for a deeper understanding of how inhibitors and facilitators evolve during the process of acquiring exercise habits, contributing new insights to the theoretical and practical development of stage-based interventions.

This study has several limitations. First, due to the cross-sectional design of this study, causal relationships between behavior change stages and the reported inhibitors or facilitators cannot be inferred. To establish temporal sequences and better understand the dynamics of exercise behavior change, future longitudinal studies are warranted. Second, researchers' subjectivity may have influenced the results, particularly in the process of categorizing inhibitors and facilitators and developing coding rules. However, to minimize this influence, we enhanced the validity of the categorization by having 2 researchers independently review the responses that included all frequently used words. Third, because this study was conducted among individuals registered with an internet panel, the generalizability of the findings may be limited. To assess potential biases, we compared the distribution of household income and behavior change stages in our sample with those reported in national datasets. Compared with the general Japanese population, respondents in this study tended to have higher household income and were more likely to be in the preparation and subsequent stages. These results suggest that the participants in our study may represent a subgroup with higher internet literacy, greater financial resources, and a stronger interest in health behaviors. Detailed comparisons with national statistics

XSL•FO
RenderX

are presented in Multimedia Appendix 8 and Multimedia Appendix 9. Fourth, all data, including the self-assessment of behavior change stages and the open-ended responses, were self-reported and therefore may be subject to recall bias or social desirability bias. Fifth, the survey was conducted during a period influenced by the COVID-19 pandemic, which may have affected respondents' exercise behaviors, perceived barriers, and available opportunities. These external factors could have shaped both the frequency and content of the reported inhibitors and facilitators.

Conclusions

These findings contribute to a deeper understanding of the dynamic mechanisms underlying behavior change, particularly in relation to exercise habit formation. By identifying how specific inhibitors and facilitators shift across behavior change stages, this study emphasizes that exercise promotion cannot be effectively approached through one-size-fits-all strategies. Instead, the results highlight the importance of tailoring interventions to individuals' readiness to change, as defined by their current stage.

For health practitioners, this suggests the need for careful stage-based assessment prior to intervention, enabling the provision of timely, stage-appropriate support—for example, motivational strategies for those in early stages, and more structural or logistical support (eg, time management, opportunity creation) for those further along the change process. In addition, insights into emerging facilitators, such as the use of personal rewards and health monitoring, can inform the development of new tools and counseling techniques to support sustained engagement.

For policymakers, the findings advocate for integrating behavioral stage frameworks into the design and implementation of public health initiatives. By incorporating simple assessments of behavioral readiness into national health checkups or workplace wellness programs, interventions can be more effectively segmented and personalized. Such approaches have the potential to improve program efficiency, maximize behavioral impact, and ultimately contribute to long-term increases in population-level physical activity and health outcomes.

Acknowledgments

We express our gratitude to all the study staff at Rakuten Insight for their commitment to data collection. We also extend our thanks to Editage for their assistance in editing and reviewing the manuscript for English language clarity.

This study was supported by the 40th Research Grant of Meiji Yasuda Life Foundation of Health and Welfare.

Disclaimer

Generative AI (ChatGPT, OpenAI) was used solely for English language editing of the manuscript. The authors are responsible for the content and interpretation of the work.

Data Availability

The datasets generated and analyzed during the current study are not publicly available due to confidentiality restrictions. If you have any questions or would like further information about the study, please do not hesitate to contact the corresponding author.

Authors' Contributions

K Sakai, KF, and MT designed this study; K Sakai collected data; K Sakai and KF performed text mining analysis and structured the analysis; K Sakai, KF, YF, SN and K Sano discussed the results; K Sakai wrote the paper. All authors revised and approved the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Frequently used word list for inhibitors with 10 or more occurrences.
[DOCX File, 27 KB - i-jmr_v14i1e77400_app1.docx]

Multimedia Appendix 2
Frequently used word list for facilitators with 10 or more occurrences.
[DOCX File, 27 KB - i-jmr_v14i1e77400_app2.docx]

Multimedia Appendix 3
Categories and behavior change stages by inhibitors and facilitators.
[DOCX File, 24 KB - i-jmr_v14i1e77400_app3.docx]

XSL•FO
RenderX

Multimedia Appendix 4
Gender-stratified distribution of inhibitor categories across behavior change stages.
[DOCX File, 284 KB - i-jmr_v14i1e77400_app4.docx]

Multimedia Appendix 5
Gender-stratified distribution of facilitator categories across behavior change stages.
[DOCX File, 306 KB - i-jmr_v14i1e77400_app5.docx]

Multimedia Appendix 6
Education-stratified distribution of inhibitor categories across behavior change stages.
[DOCX File, 751 KB - i-jmr_v14i1e77400_app6.docx]

Multimedia Appendix 7
Education-stratified distribution of facilitator categories across behavior change stages.
[DOCX File, 803 KB - i-jmr_v14i1e77400_app7.docx]

Multimedia Appendix 8
Comparison of behavior change stages distribution in this study and the national database of specific health checkups.
[DOCX File, 25 KB - i-jmr_v14i1e77400_app8.docx]

Multimedia Appendix 9
Comparison of household income distribution in this study and the comprehensive survey of living conditions.
[DOCX File, 55 KB - i-jmr_v14i1e77400_app9.docx]

Multimedia Appendix 10
The STROBE reporting checklist.
[DOCX File, 123 KB - i-jmr_v14i1e77400_app10.docx]

Multimedia Appendix 11
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File, 23 KB - i-jmr_v14i1e77400_app11.docx]

Checklist 1
STROBE checklist.
[DOCX File, 123 KB - i-jmr_v14i1e77400_app12.docx]

Checklist 2
CHERRIES checklist.
[DOCX File, 23 KB - i-jmr_v14i1e77400_app13.docx]

References

1. Smith PJ, Merwin RM. The role of exercise in management of mental health disorders: an integrative review. Annu Rev Med 2021 Jan 27;72:45-62. [doi: 10.1146/annurev-med-060619-022943] [Medline: 33256493]
2. Valenzuela PL, Ruilope LM, Santos-Lozano A, et al. Exercise benefits in cardiovascular diseases: from mechanisms to clinical implementation. Eur Heart J 2023 Jun 1;44(21):1874-1889. [doi: 10.1093/eurheartj/ehad170] [Medline: 37005351]
3. Health Japan 21 (third phase). Ministry of Health, Labour and Welfare. URL: https://www.mhlw.go.jp/stf/seisakunitsuite/bunya/kenkou_iryou/kenkou/kenkounippon21_00006.html [accessed 2025-03-01]
4. Results of a public opinion poll on the status of sports implementation. Japan Sports Agency. URL: https://www.mext.go.jp/sports/b_menu/houdou/jsa_00167.html [accessed 2025-04-07]
5. Marcus BH, Simkin LR. The transtheoretical model: applications to exercise behavior. Med Sci Sports Exerc 1994 Nov;26(11):1400-1404. [Medline: 7837962]
6. Kleis RR, Hoch MC, Hogg-Graham R, Hoch JM. The effectiveness of the transtheoretical model to improve physical activity in healthy adults: a systematic review. J Phys Act Health 2021 Jan 1;18(1):94-108. [doi: 10.1123/jpah.2020-0334] [Medline: 33260143]

XSL•FO
RenderX

7. Martins J, Costa J, Sarmento H, et al. Adolescents' perspectives on the barriers and facilitators of physical activity: an updated systematic review of qualitative studies. Int J Environ Res Public Health 2021 May 6;18(9):4954. [doi: 10.3390/ijerph18094954] [Medline: 34066596]
8. Jiménez-Zazo F, Romero-Blanco C, Castro-Lemus N, Dorado-Suárez A, Aznar S. Transtheoretical model for physical activity in older adults: systematic review. Int J Environ Res Public Health 2020 Dec 11;17(24):9262. [doi: 10.3390/ijerph17249262] [Medline: 33322327]
9. Romain AJ, Bortolon C, Gourlan M, et al. Matched or nonmatched interventions based on the transtheoretical model to promote physical activity. A meta-analysis of randomized controlled trials. J Sport Health Sci 2018 Jan;7(1):50-57. [doi: 10.1016/j.jshs.2016.10.007] [Medline: 30356472]
10. Kimura H, Asahi K, Tanaka K, et al. Health-related behavioral changes and incidence of chronic kidney disease: The Japan Specific Health Checkups (J-SHC) Study. Sci Rep 2022 Sep 29;12(1):16319. [doi: 10.1038/s41598-022-20807-2] [Medline: 36175537]
11. McPhail SM, Schippers M, Marshall AL, Waite M, Kuipers P. Perceived barriers and facilitators to increasing physical activity among people with musculoskeletal disorders: a qualitative investigation to inform intervention development. Clin Interv Aging 2014;9:2113-2122. [doi: 10.2147/CIA.S72731] [Medline: 25584023]
12. Oba T, Takano K, Katahira K, Kimura K. Revisiting the transtheoretical model for physical activity: a large-scale cross-sectional study on Japanese-speaking adults. Ann Behav Med 2024 Feb 10;58(3):167-178. [doi: 10.1093/abm/kaad069] [Medline: 38166169]
13. Planchard JH, Corrion K, Lehmann L, d'Arripe-Longueville F. Worksite physical activity barriers and facilitators: a qualitative study based on the transtheoretical model of change. Front Public Health 2018;6:326. [doi: 10.3389/fpubh.2018.00326] [Medline: 30525018]
14. Liu KT, Kueh YC, Arifin WN, Kim Y, Kuan G. Application of transtheoretical model on behavioral changes, and amount of physical activity among university's students. Front Psychol 2018;9:2402. [doi: 10.3389/fpsyg.2018.02402] [Medline: 30618907]
15. Di Gangi Samuel YP. Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. Qual Rep 2011;16:730-744. [doi: 10.46743/2160-3715/2011.1085]
16. Sakai K, Nagata T, Odagami K, Nagata M, Mori T, Mori K. Supervisors' work engagement is associated with interactivity and carefulness in supporting subordinates: a cross-sectional study using text mining in Japan. J Occup Environ Med 2023 Jul 1;65(7):e465-e471. [doi: 10.1097/JOM.0000000000002854] [Medline: 36998167]
17. Tinius R, Nagpal TS, Edens K, Duchette C, Blankenship M. Exploring beliefs about exercise among pregnant women in rural communities. J Midwifery Womens Health 2020 Jul;65(4):538-545. [doi: 10.1111/jmwh.13080] [Medline: 32277590]
18. Boutevillain L, Dupeyron A, Rouch C, Richard E, Coudeyre E. Facilitators and barriers to physical activity in people with chronic low back pain: a qualitative study. PLOS ONE 2017;12(7):e0179826. [doi: 10.1371/journal.pone.0179826] [Medline: 28742820]
19. Sakai K, Nagata T, Mori T, et al. Research topics in occupational medicine, 1990-2022: a text-mining-applied bibliometric study. Scand J Work Environ Health 2024 Oct 1;50(7):567-576. [doi: 10.5271/sjweh.4177] [Medline: 39292164]
20. Sakai K, Igarashi Y, Tounai S, et al. Key issues in Japan's public health centers to prepare for future pandemics: a text mining study using a topic model. BMC Health Serv Res 2024 May 17;24(1):636. [doi: 10.1186/s12913-024-11094-w] [Medline: 38760814]
21. Horii D, Kaneda H, Okuda E, Sugiyama Y. Quantitative analysis of qualitative data on factors affecting college students' motivation to exercise. ASSRJ 2024;11(1):188-198. [doi: 10.14738/assrj.111.16290]
22. Wang L, Guo Y, Yin X, Wang Y, Tong R. Exploring the determinants of health‐promoting behaviors among miners: a text mining and meta‐analysis. Applied Psych Health & Well 2024 Feb;16(1):3-24. [doi: 10.1111/aphw.12465]
23. Tabuchi T, Shinozaki T, Kunugita N, Nakamura M, Tsuji I. Study profile: the Japan "Society and New Tobacco" Internet Survey (JASTIS): a longitudinal internet cohort study of heat-not-burn tobacco products, electronic cigarettes, and conventional tobacco products in Japan. J Epidemiol 2019 Nov 5;29(11):444-450. [doi: 10.2188/jea.JE20180116] [Medline: 30318495]
24. Nagata T, Odagami K, Nagata M, Adi NP, Mori K. Protocol of a study to benchmark occupational health and safety in Japan: W2S-Ohpm study. Front Public Health 2023;11:1191882. [doi: 10.3389/fpubh.2023.1191882] [Medline: 38026287]
25. 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. Ann Intern Med 2007 Oct 16;147(8):573-577. [doi: 10.7326/0003-4819-147-8-200710160-00010] [Medline: 17938396]
26. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2020;6(3):e34. [doi: 10.2196/jmir.6.3.e34]
27. Ministry of Internal Affairs and Communications. Statistics Bureau. 2020. URL: https://www.e-stat.go.jp/dbview?sid=0003448237 [accessed 2025-03-02]
28. Oka K, Takenaka K, Miyazaki Y. Assessing the stages of change for exercise behavior among young adults: the relationship with self-reported physical activity and exercise behavior. Jpn Health Psychol 2000:17-23.

XSL•FO
RenderX

29. Higuchi K. New quantitative text analytical method and KH Coder software. Shakaigaku Hyoron, JSR 2017;68(3):334-350. [doi: 10.4057/jsr.68.334]
30. Sakai K, Nakazawa S, Fukai K, Furuya Y, Korenaga M, Tatemichi M. Marketing strategies for promoting workplace hepatitis B and C virus testing: a cross-sectional study using correspondence analysis in Japan. Front Public Health 2025;13:1522850. [doi: 10.3389/fpubh.2025.1522850] [Medline: 40115348]
31. Beh EJ, Lombardo R. In: Beh EJ, Lombardo R, editors. Correspondence Analysis: Theory, Practice and New Strategies: John Wiley & Sons; 2014. [doi: 10.1002/9781118762875]
32. Mbabazi J, Kanmodi KK, Kunonga E, Tolchard B, Nnyanzi LA. Barriers and facilitators of physical activity. J Health Allied Sci NU 2023 Jan;13(1):019-027. [doi: 10.1055/s-0042-1753561]
33. Zhu S, Sinha D, Kirk M, et al. Effectiveness of behavioural interventions with motivational interviewing on physical activity outcomes in adults: systematic review and meta-analysis. BMJ 2024 Jul 10;386:e078713. [doi: 10.1136/bmj-2023-078713] [Medline: 38986547]
34. Turner RR, Steed L, Quirk H, et al. Interventions for promoting habitual exercise in people living with and beyond cancer. Cochrane Database Syst Rev 2018 Sep 19;9(9):CD010192. [doi: 10.1002/14651858.CD010192.pub3] [Medline: 30229557]
35. Nicolson PJA, Bennell KL, Dobson FL, Van Ginckel A, Holden MA, Hinman RS. Interventions to increase adherence to therapeutic exercise in older adults with low back pain and/or hip/knee osteoarthritis: a systematic review and meta-analysis. Br J Sports Med 2017 May;51(10):791-799. [doi: 10.1136/bjsports-2016-096458] [Medline: 28087567]
36. Bunting JW, Withers TM, Heneghan NR, Greaves CJ. Digital interventions for promoting exercise adherence in chronic musculoskeletal pain: a systematic review and meta-analysis. Physiotherapy 2021 Jun;111:23-30. [doi: 10.1016/j.physio.2020.08.001] [Medline: 33316865]
37. Glowacki K, Duncan MJ, Gainforth H, Faulkner G. Barriers and facilitators to physical activity and exercise among adults with depression: a scoping review. Ment Health Phys Act 2017 Oct;13:108-119. [doi: 10.1016/j.mhpa.2017.10.001]
38. Korinek EV, Phatak SS, Martin CA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
39. Romain AJ, Horwath C, Bernard P. Prediction of physical activity level using processes of change from the transtheoretical model: experiential, behavioral, or an interaction effect? Am J Health Promot 2018 Jan;32(1):16-23. [doi: 10.1177/0890117116686900] [Medline: 29214837]
40. Henchoz Y, Zufferey P, So A. Stages of change, barriers, benefits, and preferences for exercise in RA patients: a cross-sectional study. Scand J Rheumatol 2013;42(2):136-145. [doi: 10.3109/03009742.2012.724707] [Medline: 23244196]
41. Lee C. Attitudes, knowledge, and stages of change: a survey of exercise patterns in older Australian women. Health Psychol 1993 Nov;12(6):476-480. [doi: 10.1037//0278-6133.12.6.476] [Medline: 8293731]
42. Mclaughlin M, Delaney T, Hall A, et al. Associations between digital health intervention engagement, physical activity, and sedentary behavior: systematic review and meta-analysis. J Med Internet Res 2021 Feb 19;23(2):e23180. [doi: 10.2196/23180] [Medline: 33605897]
43. Green EP, Whitcomb A, Kahumbura C, et al. "What is the best method of family planning for me?": a text mining analysis of messages between users and agents of a digital health service in Kenya. Gates Open Res 2019;3:1475. [doi: 10.12688/gatesopenres.12999.1] [Medline: 31410395]
44. Romain AJ, Caudroit J, Hokayem M, Bernard P. Is there something beyond stages of change in the transtheoretical model? The state of art for physical activity. Can J Behav Sci 2018;50(1):42-53. [doi: 10.1037/cbs0000093]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TTM: transtheoretical model

Edited by N Cahill, TDA Cardoso; submitted 13.05.25; peer-reviewed by H Sasai, S Narimani; revised version received 27.09.25; accepted 29.09.25; published 24.10.25.

Please cite as:
Sakai K, Fukai K, Furuya Y, Nakazawa S, Sano K, Tatemichi M
Mapping the Dynamics of Inhibitors and Facilitators of Exercise Behavior Within the Transtheoretical Model: Nationwide Cross-Sectional Study Using Text Mining Analysis
Interact J Med Res 2025;14:e77400
URL: https://www.i-jmr.org/2025/1/e77400
doi:10.2196/77400

XSL•FO
RenderX

© Kosuke Sakai, Kota Fukai, Yuko Furuya, Shoko Nakazawa, Kei Sano, Masayuki Tatemichi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 24.10.2025. 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

Long-Term Engagement of Diverse Study Cohorts in Decentralized Research: Longitudinal Analysis of "All of Us" Research Program Data

Vijay Yadav, MS; Elias Chaibub Neto, PhD; Megan Doerr, MS; Abhishek Pratap, PhD; Larsson Omberg, PhD

Sage Bionetworks, 2901 3rd Ave #330, Seattle, WA, United States

Corresponding Author:
Vijay Yadav, MS
Sage Bionetworks, 2901 3rd Ave #330, Seattle, WA, United States

Abstract

Background: The generalizability of clinical research hinges on robust study designs, which include the recruitment and maintenance of a representative study population. This study examines the evolution of the demographic characteristics of 329,038 participants who enrolled and participated in The *All of Us* Research Program (AoURP), a decentralized study aimed at representing the diversity of the United States.

Objective: The primary objectives of this study were to assess alterations in the demographic composition of the cohort at different protocol stages within AoURP, while analyzing completion rates and timeframes for survey and substudy completion. Additionally, we examined how participant interactions with the program impacted engagement and survey responses.

Methods: We conducted a longitudinal analysis of the AoURP data, tracking changes in demographic composition, completion rates, and completion times for surveys and substudies. Comparative analyses were performed to assess differences in engagement and survey completion based on sociodemographic characteristics of participants involved in postenrollment study components.

Results: The sociodemographic composition of the cohort that participated in the postenrollment study (eg, optional components) differed significantly from that of the recruited population. The proportion of self-identified White participants increased by 21.2%, whereas the proportion of Black or African American participants decreased by 12.18% (P=.02). Participants who identified as White (n=93,614, 52.7%) and NonHispanic (n=109,279, 42.21%) were more engaged compared to those identifying as Black or African American (n=10,887, 15.76%), Asian (n=4274, 38.72%), or Hispanic (n=12,530, 20.7%; P=.006). Participants' response times to study surveys and completeness varied across all demographic groups (P<.001). Furthermore, those identifying as White skipped fewer survey questions (1.19) compared to those identifying as Black or African American (1.40) or other racial and ethnic identities (P<.001).

Conclusions: The AoURP dataset serves as an exceptional resource for investigating diverse public health concerns. However, the longitudinal analysis of participant-level data underscores a significant skew in population diversity, suggesting the need for targeted strategies to enhance engagement and retention across all groups. Ensuring diversity in the cohort is essential for maintaining the study's representativeness and the broad applicability of its findings.

(Interact J Med Res 2025;14:e56803) doi:10.2196/56803

KEYWORDS

digital health; engagement behavior; All of Us Research Program; retention; decentralized research cohorts

Introduction

There is a history of lack of racial, ethnic, and gender diversity in health studies [1,2]. This lack of diversity can lead to issues related to the generalizability of research findings and equity in health care. However, lack of plurality has decreased in the last three decades, partially due to regulatory and policy efforts within government agencies, such as the National Institutes of Health [3], Food and Drug Administration [4], and Department of Health Services [5], that have sought to enhance minority participation in clinical research as well as identified scientific need [6]. As the demographic makeup of the United States is becoming more pluralistic, diverse study populations are necessary to have representative and translatable observations. Much of the effort to increase minority representation in research has focused on the recruitment of diverse study populations. However, the recruited population is not always representative of the population that is ultimately studied, due to engagement rates being different across sociodemographic groups. This has been demonstrated in remote observational studies, where there is less direct interaction with participants compared to studies which require in-person visits [7]. Poor engagement can lead to study failure due to reduced sample sizes, causing loss of power or imbalanced study populations. Prior research has

XSL•FO
RenderX

shown that remote-only studies demonstrate different engagement rates based on demographic features including disease presence or absence, age, race or ethnicity, and recruitment methods [6-8]. When the engagement rates vary across different populations, there is a risk that the effective study population (ie, the population with longitudinal data) becomes unrepresentative of the originally recruited cohort.

The *All of Us* Research Program (AoURP) was launched in May 2018, with an aim to recruit more than 1 million participants living in the United States to accelerate health research and precision medicine. AoURP has specifically focused on recruiting demographic categories that have historically been underrepresented in biomedical research (UBRs) and has largely succeeded in this objective through partnership with more than 340 recruitment sites nationwide [3]. During the first year of the AoURP, 80% of recruited participants self-identified as belonging to one or more UBR populations [9]. In this study, we explore long-term engagement within AoURP by exploring participation in optional components of the AoURP study (eg, surveys and substudies that could be performed post enrollment), the time that it takes participants to complete optional surveys, and survey response completeness. We hypothesize that this information, combined with data on specific interactions of participants with the AoURP can be used to improve and develop strategies that promote sustained engagement across the diverse demographics. This study primarily assesses changes in cohort demographics across AoURP protocol stages, completion rates for surveys and substudies, and the influence of participant interactions on engagement and responses. These insights provide actionable strategies for sustaining diverse, representative cohorts.

Methods

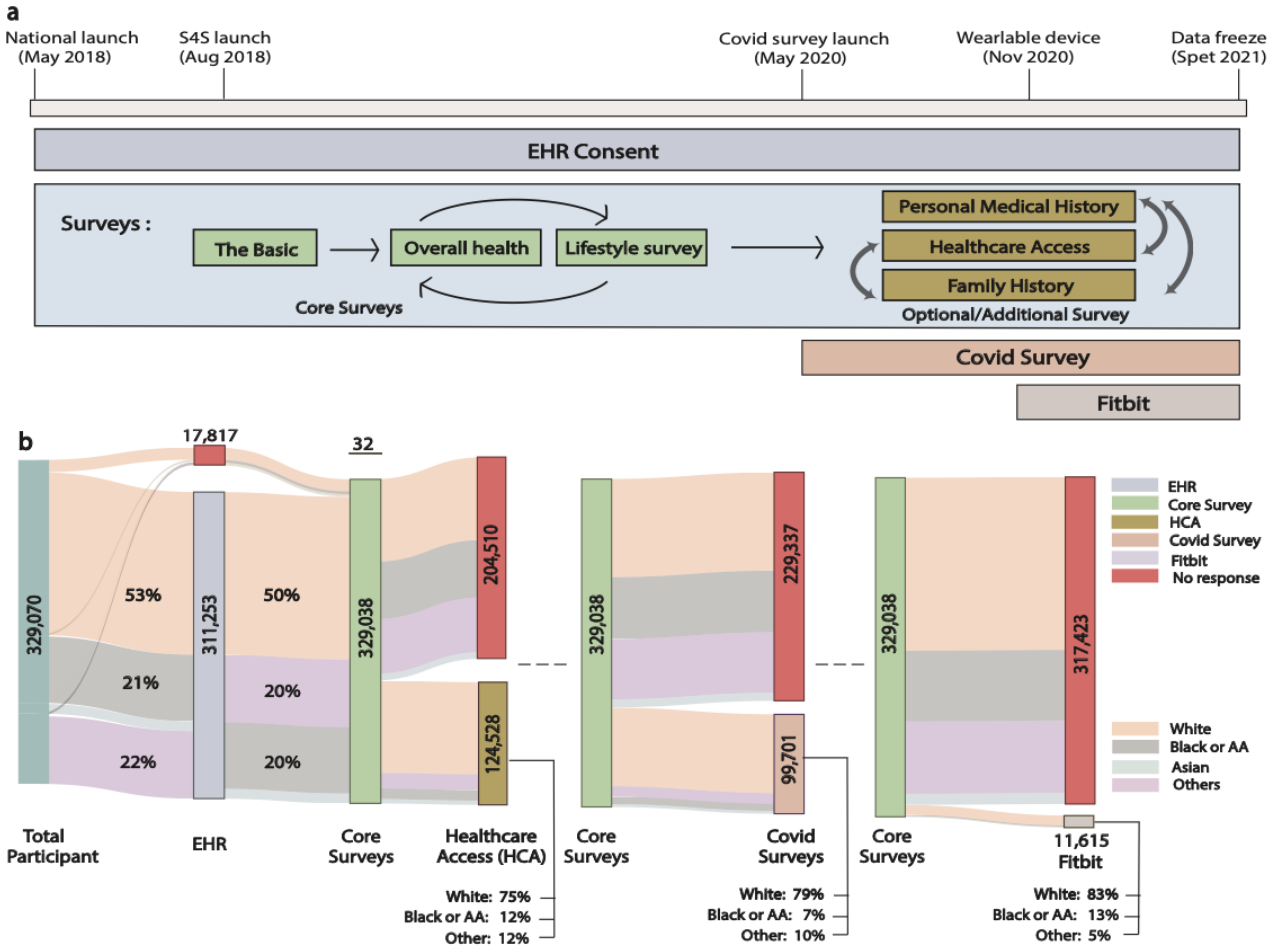
Study Design and Participants

This study is a longitudinal study that analyzes data from participants in the AoURP, encompassing individuals aged 18 and older residing in the United States, irrespective of race, ethnicity, sex, gender, or sexual orientation [10,11]. Data from the 329,038 participants in the September 2021 AoURP data freeze were analyzed. The AoURP protocol has evolved since its launch in 2018 (Figure 1A). At present, after consenting, participants are given the option to share their electronic health record (EHR) data, provide a biosample, and answer 3 core surveys (ie, The Basics, Overall Health, and Lifestyle) [12]. Participants can also respond to additional optional health surveys including Healthcare Access (HCA), Personal Medical History (PMH), and Family Medical History (FMH), at any time and in any order after enrollment (Table S1 in Multimedia Appendix 1) [12]. In addition to these components, AoURP included two additional optional substudies, at the time of analysis: (1) a COVID-19 survey and (2) a "bring your own device" (BYOD) Fitbit study, which allows participants to share data from any Fitbit wearable device (Google Inc, Mountain View, CA) owned by them. The AoURP also collects additional data, such as physical measurements; however, our analysis is restricted to recruitment and engagement characteristics based on participant-provided information (eg, health and demographic surveys) and BYOD Fitbit data.

The core surveys included information about the participants' self-reported demographics, which was used to explore differences in participation. Specifically, we used responses to the questions on (1) date of birth; (2) What was your biological sex assigned at birth; (3) Which categories describe you (race); (4) Which categories describe you (ethnicity); and (5) What is your annual household income (from all sources) [12]. Based on the *All of Us* (AoU) Researcher Workbench, we categorized age into three bands: younger (18 - 44 years), middle aged (45 - 64 years), and older adults (≥65), and household income into 4 bands (<$50,000, $50,000-$100,000, $100,000-$200,000, >$200,000). For other multiple-choice questions, we included categories with at least 1.8% respondents. This resulted in the following categories: male and female for biological sex at birth; White, Black or African American, and Asian for race; and NonHispanic and Hispanic for ethnicity. These categories were used to assess changes in cohort demographics during the course of the study, with the primary outcomes under investigation being (1) engagement (ie, number of participants who remain engaged by completing optional components after enrollment), (2) response time (ie, time between invitation and completion of optional components), and (3) completeness of response (ie, how often participants chose to answer specific questions). Understanding these changes is essential for examining shifts in participant demographics and engagement over time, which helps identify trends and biases, ensuring the cohort representativeness and development of targeted strategies for improved engagement and retention.

XSL•FO
RenderX

Figure 1. Protocol and changes in engagement across components in study. (A) Timeline for components of protocol and journey of a participant in the AoU program. Based on respective protocol launch dates, participants had the option to consent to share EHR data, participate in COVID-19 surveys, and share data from a BYOD Fitbit data. Each survey block reflects the ordering of core and additional surveys configured in the program. (B) Sankey diagram showing the population distribution within the AoURP throughout the protocol journey. The nodes represent different types of events (eg, EHR consent, completion of optional surveys, BYOD Fitbit data sharing). Flows between notes are color-coded by self-identified race (ie, White, Black, or African American, Asian, or other). The width of the nodes and links provide quantitative information. AA: African American, AoURP: *All of Us* Research Program; BYOD: bring your own device; EHR: electronic health records; HCA: Healthcare Access survey.



Data Processing

For the primary analysis, we built an engagement cohort consisting of participants who successfully enrolled in the study and responded to the demographic questions using the AoU Researcher Workbench [13]. Since all components of the study were not launched at the same time (eg, BYOD Fitbit and the sharing of electronic health records using Sync4Science [14] were added later; see Figure 1A for the timeline), we conducted a secondary analysis using enrollment dates extracted from the observation table to evaluate the effect of timing of enrollment on these components. Specifically, this secondary analysis included individuals who enrolled after Sync4Science launch (August, 2018) and after the BYOD Fitbit protocol being launched (November, 2020). For all cohorts, activities where participants completed the surveys were considered as "active/engaged" events. All survey questions included a "prefer not to answer" option. Survey completeness was analyzed across participants by counting the number of questions answered as "prefer not to answer".

Statistical Analysis

Various statistical techniques were used for the three different analyses. Whenever possible we favored the use of nonparametric statistical tests and resampling techniques to avoid relying on distributional assumptions. These included the Kruskal-Wallis test (instead of analysis of variance), Mann-Whitney tests for pairwise group comparisons), bootstrapping for computation of confidence intervals, and permutation tests for calculating statistical significance. For all analyses, a significance level of $P<.05$ was used with Bonferroni correction for multiple testing [15].

To analyze engagement across self-reported age, sex at birth, race, household income, and ethnicity groups, we compared the proportion of participants responding in each demographic group between the core survey and optional components using a χ^2 test for homogeneity.

Participants' response time, as measured by the number of days between joining the study and completing (or joining) an optional component, was analyzed using the nonparametric Kruskal-Wallis one-way analysis of variance for each optional

XSL•FO
RenderX

component separately, as the data was not normally distributed [16]. To further understand the variability between each group, pairwise differences between groups were analyzed using the Python language (version number 3.9; Python Software Foundation), including the *statannotation* package (version 0.4.2) [17], along with the with Mann-Whitney integrated statistical test and Bonferroni for multiple testing correction [18,19].

To better understand the relationship between various self-reported demographic variables, a linear mixed model (LMM) was used to analyze the response time across the 3 surveys concomitantly. To improve the fit of the model, response time data was log-transformed before feeding into the model. All self-reported demographic variables were fitted into the model to evaluate the effect produced by each predictor. Age was treated as a continuous variable, while other variables were fitted as categorical variables. The coefficient of relative change for each variable was reported by the LMM model, and the log-transformed relative change between each group was back-transformed to a percentage change in response time across groups for all the categories, using formula $100(e^{\text{coefficient}}-1)$. To further evaluate confidence intervals, bootstrapping was used. To account for the repeated measurements of each subject, the bootstrap confidence intervals for the LMMs were constructed based on 1000 iterations of the following couple of steps. First, we obtained a bootstrap sample from the original data by sampling with replacement at the subject level (ie, assuming that our dataset contains data from *N* subjects, we grouped the repeated measurements of each subject as a "subject block," and then sampled with replacement N "subject blocks" out of the N blocks in the original dataset.) Second, for each bootstrap sample, we fit a LMM and computed the coefficient of relative change of each variable as described above. Finally, we computed 95% CI for the coefficients of relative change using the percentile interval method, where the lower and upper confidence bounds correspond to the 0.025 and 0.975 quantiles of the distribution of coefficients generated across the 1000 bootstrap iterations.

We independently examined outliers in the response time for optional surveys to check if these outliers were the cause of any systematic differences. Outliers comprised any individual who had a response time less than the first quartile (Q1)–1.5×IQR or greater than Q3+1.5×IQR. To evaluate the differences, we computed the proportion of outlier participants for each demographic group and then conducted a χ^2 test (test of homogeneity) for testing the equality of outlier participants proportion across different demographic groups.

Completeness of response was evaluated by counting the number of questions where a participant responded "prefer not to answer." First, the nonparametric Kruskal-Wallis test was used to evaluate variability across different groups (since the data was not normally distributed). However, due to ties in the completeness of the response variable, we adopted a permutation test to assess differences between the groups [12]. The permutation test was conducted by comparing the observed value of the ANOVA F-statistic computed in the original data (F_{obs}) against the permutation null distribution of the F-statistic computed on permuted versions of the data (F*), where the group labels were randomly shuffled. The test was based on 1000 data permutations and the permutation *P* value was computed as the proportion of times the permuted F-statistic was greater or equal to the observed test statistic (ie, $P=(1+\text{sum}\{i=1..B \text{ indicator}\{F^*>=F_{obs}\}\})/(1+B)$, for B=1000). Due to the large sample set, it is possible that we may be detecting very small effect sizes with very high statistical significance across different categories. To assess the effect size, 95% CI were computed using nonparametric bootstrapping for the pairwise group mean difference for each group [20].

Ethical Considerations

This study used deidentified data from the *All of Us* Research Program, which obtained institutional review board (IRB) approval and informed consent from participants for secondary research [21]. The original data collection was approved by the AoURP IRB [21], and no additional IRB approval was required for this secondary analysis. According to the All of Us Responsible Conduct of Research training, analyses using deidentified data on the Researcher Workbench do not require a separate IRB review [22], as the research involves no direct interaction with participants. All analyses adhered to ethical principles, ensuring privacy and confidentiality as outlined by the AoURP.

Results

Population Characteristics

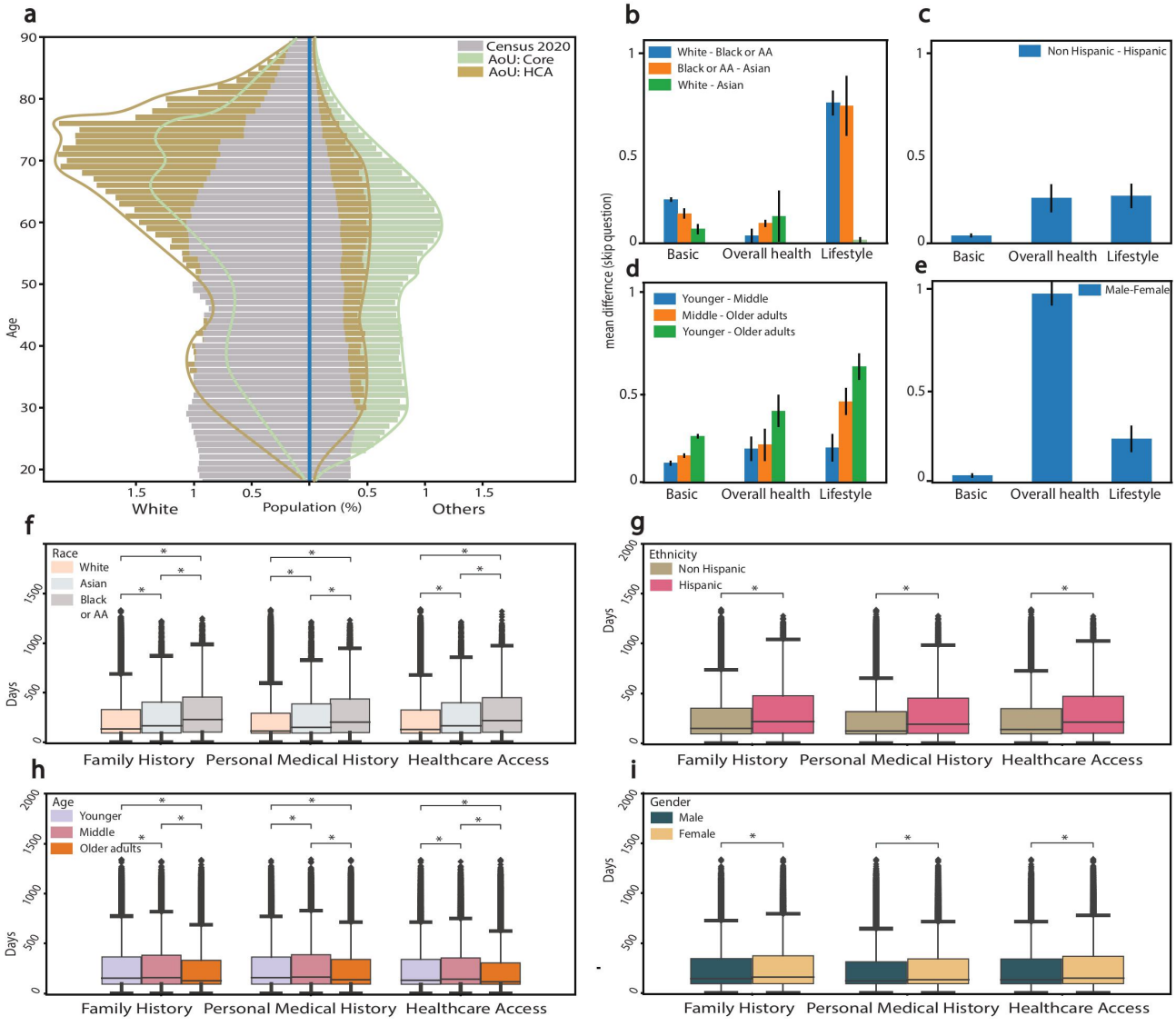
Data from 329,078 consented participants were made available as part of the version 5 data release [23] by AoURP. Of the total consented participants, 329,038 (99.98%) completed all core surveys and were considered successfully enrolled in the program. Additionally, 311,253 participants (94.59%) consented to share all their electronic health records; including retrospective medical records from before the AoURP study was launched. Post completion of core surveys, participants had the option to complete three optional health surveys. A significantly smaller proportion completed these three optional surveys: (1) HCA-37.84% (n=124,528/329,078), (2) FMH-117,693 (35.76%), and (3) PMH-113,830 (34.59%) (Figure 1B). Both the COVID survey and the BYOD Fitbit data collection were launched at later stages of the study (Figure 1A). A total of 99,701 (30.3%) people participated in the first COVID survey and 11,615 (3.52%) participated in BYOD Fitbit data sharing. Participants who enrolled before the BYOD Fitbit data collection protocol was launched were less likely to consent to share their BYOD Fitbit data compared to those who enrolled after the protocol launch (3.51% vs 6.51%, respectively).

The AoURP aims to recruit persons from the UBR populations, and recruitment was successful, with 24.36% (n=80,154) of the participants responding to the core surveys with a self-reported race of Black or African American and Asian, while 53.98% (n=177,615) self-reported as White. We compared the distribution of ages and race to the 2020 US Census data [24], which showed the enrichment of self-identified races other than White, as well as middle-aged individuals, in the AoU cohort (Figure 2A) at baseline. Additionally, more participants reported

XSL•FO
RenderX

being female (Figure S1a in Multimedia Appendix 1) and Hispanic ethnicity (Figure S1b in Multimedia Appendix 1) compared to the census.(However, this initial enrichment of people self-identifying as belonging to one or more UBR populations diminished as participants engaged with optional components of the protocol. Distributions across demographic variables for each component of the study is shown in Table 1. Moreover, a total of 1416 (0.4%) deaths were reported by the end of version 5 data release.

Figure 2. (a) Population pyramids divided by self-reported race (White vs other) see Multimedia Appendix 1 for additional populations pyramids. Three different pyramids are displayed for comparison for three different populations representing: those who responded to: AoURP core survey, one optional survey (HCA) and the 2020 US census. At baseline, based on response to the core survey) cohorts tend to be older and less white compared to the 2020 US Census, while responses to later optional surveys showed that the proportion of self-reported White participants increased significantly and a sudden drop in participants self-identifying from other racial groups. (B-E) represents the mean difference of number of skipped questions between demographic groups specific to each survey for: (b) *Race*, (c)*: Ethnicity*, (d)*: Age*, (e)*: Sex at birth*. The bar plot with the error bar represents the bootstrap (1000 iterations) 95% confidence interval for pairwise group mean differences. (f): Race, (g): Ethnicity, (h): Age and (i): Sex at birth) represent pairwise statistically significant differences (P<.001) across all the self-reported demographic groups of the participants responded to the additional surveys. AA: African American; AoU: All of Us. HCA: Health Care Access. *P<.001.



XSL•FO
RenderX

Table . Demographics characteristics and percentage distribution of participants in each component.

Groups	Core survey[a] (n=329,038), n (%)	HCA[b] survey (n=124,528), n (%)	Family history (n=117,693), n (%)	Medical history (n=113,830), n (%)	EHR[c] (n=271,421), n (%)	Covid survey (n=99,701), n (%)	Fitbit (n=11,615), n (%)
Age (years)							
18 - 44	108,254 (32.9)	35,740 (28.7)	33,307 (28.3)	32,328 (28.4)	88,212 (32.5)	23,130 (23.2)	3,984 (34.3)
45 - 64	121,415 (36.9)	42,090 (33.8)	39,545 (33.6)	38,019 (33.4)	100,697 (37.1)	33,698 (33.8)	4,321 (37.2)
≥65	99,369 (30.2)	46,698 (37.5)	44,841 (38.1)	43,483 (38.2)	82,512 (30.4)	42,971 (43.0)	3310 (28.6)
Sex							
Male	125,034 (38)	42,090 (33.8)	39,663 (33.7)	38,475 (33.8)	104,497 (38.5)	33,699 (33.8)	3,357 (28.9)
Female	199,726 (60.7)	81,441 (65.4)	77,207 (65.6)	74,559 (65.5)	163,395 (60.2)	65,304 (65.5)	8,200 (70.6)
Race							
White	177,681 (54)	93,645 (75.2)	89,094 (75.7)	86,625 (76.1)	144,124 (53.1)	78,963 (79.2)	9,698 (83.5)
African American	69,098 (21)	10,958 (8.8)	10,004 (8.5)	9,448 (8.3)	58,898 (21.7)	7,278 (7.3)	581 (5)
Asian	11,187 (3.4)	4,234 (3.4)	4,002 (3.4)	3,870 (3.4)	8,685 (3.2)	2,892 (2.9)	372 (3.2)
Ethnicity							
NonHispanic	258,953 (78.7)	109,336 (87.8)	103,570 (88)	100,398 (88.2)	212,523 (78.3)	89,631 (89.9)	10,674 (91.9)
Hispanic	60,543 (18.4)	12,577 (10.1)	11,652 (9.9)	11,041 (9.7)	51,027 (18.8)	7,876 (7.9)	767 (6.6)
Income (USD)							
<50,000	139,183 (42.3)	38,106 (30.6)	35,661 (30.3)	34,035 (29.9)	116,982 (43.1)	27,916 (28)	2,532 (21.8)
50,000-100,000	59,556 (18.1)	32,626 (26.2)	31,071 (26.4)	30,279 (26.6)	48,041 (17.7)	27,518 (27.6)	3,554 (30.6)
100,000-200,000	46,394 (14.1)	28,766 (23.1)	27,422 (23.3)	26,750 (23.5)	36,913 (13.6)	24,526 (24.6)	3,520 (30.3)
>200,000	20,071 (6.1)	12,702 (10.2)	12,005 (10.2)	11,724 (10.3)	15,851 (5.84)	10,469 (10.5)	1,371 (11.8)

[a]Core Survey at enrollment.

[b]HCA: Health Care Access.

[c]EHR: electronic health records.

Differences in Engagement

The relative shift in the proportions of participants is reported as percentage changes. We observed an increase of 21.2% in the proportion of self-identified White participants and correspondingly a decrease of Black or African American respondents by 12.18% in the HCA compared to the Core surveys. Similarly, the proportion of participants who identified as NonHispanic (of all races) increased by 9.1%, to 87.8% (n=109,336) of the total respondent pool in the HCA survey, while people self-identifying as Hispanic decreased by 8.33% to 10.1% of the respondent pool (Table 1 and Figure 2A). To explore the engagement further, we compared the percentage of each self-identified demographic group that continued in each optional component of the study (ie, optional surveys, EHR consent, COVID Surveys, and BYOD Fitbit) and observed that the engagement varied across most components except the BYOD Fitbit (Table 2). The engagement with optional surveys

XSL•FO
RenderX

was 3 times higher among people who identified as White, with 52.7% (n=93,614) of White participants who enrolled and also completed the optional surveys (HCA), compared to Black or African American participants (n=10,887, 15.8%), and nearly twice as high as those identifying as Asian (n=12531, 38.7%) (*P*<.001). Similarly, 42.2% (n=109,279) of those self-identifying as nonHispanic (of all races) engaged with optional surveys as compared to 20.7% (n=12,531) of self-identified Hispanic participants (*P*<.006). 63.4% (n=12,681) of participants who self-identified with high household income engaged in the optional surveys compared to 27.4% (n=38,190) of self-identified with low household income (*P*<.001). A total of 47% (n=46,743) of older adults (aged >65 years) engaged with optional surveys compared to 34.64% (n=37,504) and 33.05% (n=40,096)of the middle-aged and younger groups, respectively. Engagement was 40.8% (n=81,460) for participants who self-identified as female at birth engaged in the optional surveys compared to 33.66% (n=42,137) of self-identified male at birth. However, neither self-reported age nor sex at birth were statistically significant indicators of engagement using the χ^2 test.

The first COVID survey showed large differences in participation across age (*P*=.02), race (*P*<.001), household income (*P*<.001), and ethnicity (*P*<.001), but not by sex at birth (*P*=.45), Older individuals participated at higher rates, with 43.2% (n=42,964) of self-reported older adults compared to 27.8% (30,099) of middle-aged and 21.3% (n=25,841) of younger participants completing the first survey. Self-reported White participants (n=79,048, 44.5%) participated at more than 4 times higher levels than self-identified Black or African American participants (n=7281, 10.54%) and nearly twice those identifying as Asian (n=2880, 26.09%). Self-identified high-income participants (n=10,485, 52.4%) participated at more than approximately 2.5 times higher levels than self-identified low household income participants (n=27,876, 20%). Similarly, participants self-identifying as nonHispanic (n=89,552, 34.59%) engaged at almost 3 times higher rates than those identifying as Hispanic (n=7876, 13.01%).

We did not observe any statistically significant differences in engagement by self-reported age, race, ethnicity, or sex in participation rates for the BYOD Fitbit substudy or the consent to share EHR data.

XSL•FO
RenderX

Table . Demographics characteristics and response rate proportion specific to each survey in comparison to core surveys, (ie, response rate percentage = Number of participant responded to optional surveys/Number of participants responded to core survey)×100.

Groups	Core[a] (n)	HCA[b] survey, (%)	*P* value	Family history (%)	*P* value	Medical history, n (%)	*P* value	EHR[c], n (%)	*P* value	Covid survey, n (%)	*P* value	Fitbit[d], n (%)	*P* value
Age groups (years)			.22		.18		.19		.94		.02		.99
18 - 44	121,317	40,156 (33.1)		37,244 (30.7)		36,274 (29.9)		96,932 (79.9)		25,841 (21.3)		4,489 (3.7)	
45 - 64	108,268	37,461 (34.6)		35,295 (32.6)		33,996 (31.4)		91,162 (84.2)		30,099 (27.8)		3,898 (3.6)	
>65	99,453	46,743 (47)		44,853 (45.1)		43,461 (43.7)		82,745 (83.2)		42,964 (43.2)		3,282 (3.3)	
Sex			.41		.41		.42		.81		.45		.59
Male	125,184	42,187 (33.7)		39,683 (31.7)		38,431 (30.7)		105,530 (84.3)		33,674 (26.9)		3,380 (2.7)	
Female	199,658	81,460 (40.8)		77,068 (38.6)		74,672 (37.4)		162,122 (81.2)		65,288 (32.7)		8,186 (4.1)	
Race			<.001		<.001		<.001		0.40		<.001		.18
White	177,637	93,615 (52.7)		89,174 (50.2)		86,509 (48.7)		135,359 (76.2)		79,048 (44.5)		9,770 (5.5)	
African American	69,084	10,915 (15.8)		10,017 (14.5)		9,395 (13.6)		63,557 (92)		7,254 (10.5)		553 (0.8)	
Asian	11,040	4,272 (38.7)		3,974 (36)		3,875 (35.1)		8,578 (77.7)		2,881 (26.1)		364 (3.3)	
Ethnicity			.006		.006		.006		.41		.001		.23
NonHispanic	258,895	109,254 (42.2)		103,558 (40)		100,451 (38.8)		208,152 (80.4)		89,578 (34.6)		10,615 (4.1)	
Hispanic	60,535	12,531 (20.7)		11,623 (19.2)		11,078 (18.3)		55,147 (91.1)		7,870 (13)		787 (1.3)	
Income (USD)			.001		.001		.001		.98		<.001		.31
<50,000	139,380	38,162 (27.4)		35,639 (25.6)		34,092 (24.5)		134,125 (96.2)		27,946 (20.1)		2,537 (1.8)	
50,000-100,000	59,563	32,700 (54.9)		31,134 (52.3)		30,222 (50.7)		55,078 (92.5)		27,554 (46.3)		3,556 (6)	
100,000-200,000	46,381	28,728 (61.9)		27,411 (59.1)		26,729 (57.6)		42,346 (91.3)		24,512 (52.9)		3,520 (7.6)	
>200,000	20,008	12,685 (63.4)		12,043 (60.2)		11,785 (58.9)		18,181 (90.9)		10,494 (52.5)		1,377 (6.9)	

[a]Core survey at enrollment.

[b]HCA: Health Care Access.

[c]EHR: electronic health records.

[d]Fitbit was launched later in the study (see Figure 1). For a comparison of engagement rates between participants who joined before and after the launch of the Fitbit component (see Table S5 in Multimedia Appendix 1).

Response Time to Optional Surveys

In addition to differences in engagement, we also analyzed how long it took for participants to participate in optional surveys as measured by the time in days between enrollment and completion of the optional component. Given that only the three optional surveys (ie, HCA, PMH, and FMH) were available,

XSL•FO
RenderX

and the COVID Surveys and BYOD Fitbit, were unavailable at the launch of the study, we were only able to explore the delay for these surveys. We found differences in the response time based on self-reported age, sex at birth, race, ethnicity, and household income (all *Ps*<.001 based on Kruskal-Wallis one way analysis of variance). These results are summarized in Table S2 in Multimedia Appendix 1. Overall for the optional surveys, response times ranged from 0 to 1342 days with a mean of 247.9 (SD 233) days and median value was 141 (IQR 314.55) days. Self-reported older adults (median 117, IQR 213) took less time to respond to optional surveys compared to younger (median 135, IQR 251) and middle aged (median 138, IQR259) participants. People who identified as male at birth (median 120, IQR 221) took less time compared to people who self-identified as female at birth (median 133, IQR249). People who self-identified as White completed surveys sooner after enrollment (median 115, IQR 202) in comparison to those who self-identified as Black or African American (median 202, IQR 340), and Asian (median 153, IQR 295). People who identified as nonHispanic of any race (median 123, IQR 225) responded sooner to optional surveys as compared to people who self-identified as Hispanic (of all races) (median 188, IQR 335). People self-identified with high household income (median 119, IQR 195) took less time compared to people with low household income (median 150, IQR 279). We performed a secondary analysis to determine pairwise differences using the nonparametric Mann-Whitney test to illustrate which groups had quicker participation (Figure 2F–I). The linear mixed effect results presented in Table 3 provided valuable insights into the relationships between various demographic variables and their impact on response time. The table displays the estimated effects in terms of coefficients and relative percentage changes, with the use of bootstrap for accuracy. When examining age groups, individuals aged 45-64 years exhibited a 3.04% increase in response time compared to the reference group (18-44 years), while those aged ≥65 years displayed a decrease of −2.76%. Gender differences were also notable, with women showing a 5.65% increase in response time compared to men. In terms of race, African Americans exhibited a substantial 33.5% increase in response time compared to White individuals, while Asians and other racial groups also displayed significant differences. Ethnicity played a role as well, with Hispanics experiencing an 18.41% increase (Table 3) . Finally, income levels showed a consistent trend, with higher income brackets associated with decreased response times, highlighting the influence of socioeconomic factors on response behavior. These results offer important insights into the complex interplay between demographics and response times, shedding light on potential areas for further investigation and intervention. We conducted a secondary analysis to check if outliers were responsible for any significant difference and found no difference in the demographics. Distribution across demographic data is shown in Table S3 in Multimedia Appendix 1. We also explored the interaction effect (see Table S4 in Multimedia Appendix 1) between other demographic features and household income. The study found that the effect of age, gender, race, and ethnicity on response time was moderated by income. The response time was slower for people with higher income in all age groups, except 18-44 years. The difference was most pronounced for people aged ≥65 years. The interaction effect between gender and income was only significant for people who identify as neither male nor female. The interaction effect between race and income was significant for all race groups except White. The interaction effect between ethnicity and income was only significant for people who were not in plurality.

XSL•FO
RenderX

Table . Estimated effect (coefficient and relative % change using bootstrap) on response time.

Variables	Model	
	Coefficient	Percent change (95% CI)
Age (years)		
18-44	0	0
45-64	0.03	3.04 (2.99-3.09)
>65	–0.028	–2.76 (–2.78 to –2.69)
Gender		
Male	0	0
Female	0.055	5.65 (5.55-5.65)
Race		
White	0	0
African American	0.289	33.50 (33.40-33.57)
Asian	0.169	18.41 (18.30-18.54)
Ethnicity		
NonHispanic	0	0
Hispanic	0.169	18.41 (18.26-18.49)
Income (USD)		
<50,000	0	0
50,000-100,000	–0.045	–4.79 (–4.91 to –4.56)
100,000-200,000	–0.066	–6.39 (–6.45 to –6.31)
>200,000	–0.05	–4.88 (–4.96 to –4.78)

Completeness of Responses

To evaluate the frequency of skipped questions in each survey (survey completeness), we counted skipped questions and compared those across self-reported demographic groups (Table S5 in Multimedia Appendix 1). Skipped questions were relatively rare, with approximately 124,000 skipped questions out of 34 million answered; on average 329,038 participants skipped a mean of 1.32 (SD 0.74) questions per survey. We reported the results of "completeness of response" based on core surveys only, given that all the questions in the optional surveys were already answered except for one participant, who skipped one question. Using a nonparametric test, we observed differences in completeness for all self-reported demographic groups in the three core surveys. Mean values are reported for missing data (ie, skipped questions). Middle-aged participants skipped more questions (1.72 questions) compared to 1.54 for younger and 1.6 for seniors (P<.001). Participants who identified as female at birth skipped fewer questions (1.58) compared to participants who identified male at birth (1.99) (P<.001). Participants who self-reported as Black or African American race (1.86) and people self-reporting nonHispanic ethnicity of any race (1.71) skipped more questions compared to people of self-identified White (1.62) and Asian (1.62) race, and those who self-reported Hispanic (1.57) ethnicity of any race (P<.001). Participants with high household income skipped fewer questions (1.58) compared to participants with low household income (1.61) (P<.001). Also, we conducted pairwise group mean differences to further estimate the effect size which returned significant results (Figure 2B-E).

Secondary Analysis on Effect of Timing of Enrollment on Engagement

As all protocols were not launched at the same time, time of enrollment may have affected the engagement behavior of the participants who joined before and after the launch of the specific study components. We conducted a secondary analysis to evaluate if there were any differences in engagement between individuals who enrolled before and after the launch of BYOD Fitbit and Sync4Science separately. We found no significant differences in the demographic groups. Distribution across demographic variables for pre- and postlaunch of BYOD Fitbit (Table S6 in Multimedia Appendix 1), and for Sync4Science is shown in Table S7 in Multimedia Appendix 1.

Discussion

Principal Finding

The utility and translatability of studies is dependent on the study being carried out on a representative population. The AoURP has enrolled an extremely large cohort that aims to be representative of the diversity of the broader US population [25]. However, the data collected over time is not as representative as the enrolled population, with differences in engagement across demographic groups. This pattern is observed in both, the demographics that remained engaged and the degree

XSL•FO
RenderX

to which participants completed optional substudy components and how quickly they responded [26]. These variations are driven by socioeconomic, cultural, and structural factors, underscoring the need for tailored strategies to mitigate participation disparities [27]. Addressing these disparities is essential to maintain the cohort's representativeness, which directly impacts the validity and generalizability of the findings [28].

Comparison to Prior Work

This study builds on prior research by highlighting demographic engagement disparities in longitudinal cohort studies, an issue that has been well-documented in other population-based research efforts [29,30]. Previous studies have reported similar challenges, such as the underrepresentation of minority groups and the overrepresentation of higher socioeconomic status in voluntary participation metrics [31]. However, unlike earlier research, this study leverages the unique scale and diversity of the AoURP cohort to provide deeper insights into how engagement patterns vary over time and across demographic groups. By emphasizing socioeconomic and structural drivers of engagement disparities, this work highlights the need for target interventions necessary for retaining participants and ensuring equitable representation [32].

Strengths and Limitations

Our study benefits from a large, diverse cohort and longitudinal data, providing valuable insights into participation patterns. However, there are limitations that must be acknowledged. First, there may be biases due to self-selection and varying dropout rates among demographic groups. Overrepresentation of more engaged participants with higher socioeconomic status in optional surveys could skew health outcome estimates [7]. Second, the high dropout rates observed among underrepresented groups over time pose a challenge to maintaining sample diversity, which could limit the generalizability of findings to the broader US population [33]. Third, reliance on self-reported data, which may introduce reporting biases that affect the accuracy of the findings [27]. To address these limitations, efforts should focus on implementing strategies such as over-recruiting populations with higher attrition rates, developing tailored retention programs, and validating self-reported data through objective measures such as linked medical records [29]. These steps can enhance the reliability and applicability of the study's findings, ensuring their broader impact.

Future Directions

Based on our findings, decentralized studies such as AoURP should prioritize interventions that address engagement disparities. Over-recruiting populations with higher attrition rates and implementing targeted outreach strategies for underrepresented groups can enhance participation. Importantly, these strategies should be codesigned with input from participants to address specific barriers effectively [30]. Addressing challenges such as the digital divide, cultural differences, and logistical barriers is critical to fostering equity in research participation [31,34]. Innovative approaches such as bidirectional engagement where participants actively contribute to study design and receive meaningful feedback could further strengthen retention. Additionally, leveraging digital health tools and adaptive technologies can facilitate participant engagement while addressing socioeconomic barriers [35].

Conclusion

This study underscores the importance of maintaining representativeness in large, decentralized cohort studies such as AoURP. By identifying participation disparities and their underlying drivers, we provide actionable insights to improve cohort retention and engagement strategies. Future research should continue to explore and validate these interventions to ensure equitable and impactful scientific discoveries that benefit all segments of the population.

Acknowledgments

Research reported here was supported by the Office of the Director of the National Institutes of Health under award number U24OD023176. This work would not be possible without the partnership of the participants in the AoURP. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Data Availability

Data are available only to researchers through the All of Us cloud-based Research Workbench. The complete code used for the analysis is available through a GitHub code repository [36].

Conflicts of Interest

All authors listed are employed at Sage Bionetworks. AP is an employee of Boehringer Ingelheim, USA.

Multimedia Appendix 1
Demographic characteristics.
[DOCX File, 199 KB - i-jmr_v14i1e56803_app1.docx]

References

XSL•FO
RenderX

1. Ford JG, Howerton MW, Lai GY, et al. Barriers to recruiting underrepresented populations to cancer clinical trials: a systematic review. Cancer 2008 Jan 15;112(2):228-242. [doi: 10.1002/cncr.23157] [Medline: 18008363]
2. Bolen S, Tilburt J, Baffi C, et al. Defining "success" in recruitment of underrepresented populations to cancer clinical trials: moving toward a more consistent approach. Cancer 2006 Mar 15;106(6):1197-1204. [doi: 10.1002/cncr.21745] [Medline: 16453333]
3. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research. NIH Grants and Funding. 2023 Oct 30. URL: http://grants.nih.gov/grants/funding/women_min/guidelines_amended_10_2001.htm [accessed 2025-02-28]
4. Food and drug administration modernization act (FDAMA) of 1997. US Food and Drug Administration. 2018 Nov 3. URL: https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/food-and-drug-administration-modernization-act-fdama-1997 [accessed 2023-10-30]
5. Medicare clinical trial policies. Centers for Medicare & Medicaid Services. 2023 Oct 30. URL: https://www.cms.gov/medicare/coverage/approved-facilities-trials-registries/clinical-policies?redirect=/ClinicalTrialPolicies [accessed 2025-02-28]
6. Yancey AK, Ortega AN, Kumanyika SK. Effective recruitment and retention of minority research participants. Annu Rev Public Health 2006;27:1-28. [doi: 10.1146/annurev.publhealth.27.021405.102113] [Medline: 16533107]
7. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020;3:21. [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
8. Kolovson S, Pratap A, Duffy J, Allred R, Munson SA, Areán PA. Understanding participant needs for engagement and attitudes towards passive sensing in remote digital health studies. Int Conf Pervasive Comput Technol Healthc 2020 May;2020:347-362. [doi: 10.1145/3421937.3422025] [Medline: 33717638]
9. The All of Us Research Program Investigators. The "All of Us" Research Program. N Engl J Med 2019:668-676. [doi: 10.1056/NEJMsr1809937]
10. Who can join? All of Us Research Program. 2024 Jul 15. URL: https://www.joinallofus.org/who-can-join [accessed 2025-02-28]
11. All of us research program protocol summary. All of Us Research Program. 2024 Jul 15. URL: https://allofus.nih.gov/sites/default/files/All_of_Us_Protocol_Overview_Mar_2022.pdf [accessed 2025-02-28]
12. Survey explorer. All of Us Research Hub. 2022 Apr 25. URL: https://www.researchallofus.org/data-tools/survey-explorer/ [accessed 2025-02-28]
13. Researcher workbench. All of Us Research Hub. 2022 Feb 14. URL: https://www.researchallofus.org/data-tools/workbench/ [accessed 2025-02-28]
14. Helping patients share EHR data with researchers. Sync For Science. 2024 Apr 25. URL: http://syncfor.science/ [accessed 2025-02-28]
15. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. BMJ 1995 Jan 21;310(6973):170. [doi: 10.1136/bmj.310.6973.170] [Medline: 7833759]
16. Kruskal WH, Wallis WA. Use of ranks in one-criterion variance analysis. J Am Stat Assoc 1952 Dec;47(260):583-621. [doi: 10.1080/01621459.1952.10483441]
17. Statannotations 0.7.1. PYPI. URL: https://pypi.org/project/statannotations/ [accessed 2025-02-28]
18. Mann HB, Whitney DR. On a test of whether one of two random variables is stochastically larger than the other. Ann Math Statist 1947 Mar;18(1):50-60. [doi: 10.1214/aoms/1177730491]
19. Wilcoxon F. Individual comparisons by ranking methods. Biometrics Bulletin 1945 Dec;1(6):80. [doi: 10.2307/3001968]
20. Efron B, Tibshirani RJ. An introduction to the bootstrap: CRC Press; 1994.
21. All of Us institutional review board (IRB). All of Us Research Program | NIH. 2022 Apr 8. URL: https://allofus.nih.gov/about/who-we-are/institutional-review-board-irb-of-all-of-us-research-program [accessed 2025-11-25]
22. Do I need my project reviewed by the all of us institutional review board (IRB) in order to access this data using the researcher workbench? All of Us Research Hub. 2024 Nov 25. URL: https://www.researchallofus.org/faq/do-i-need-my-project-reviewed-by-the-all-of-us-institutional-review-board-irb-in-order-to-access-this-data-using-the-researcher-workbench/ [accessed 2025-02-25]
23. Release notes. All of Us Research Program. URL: https://support.researchallofus.org/hc/en-us/sections/360006773231-Release-Notes [accessed 2025-02-25]
24. Data: try out our new way to explore data. US census bureau. 2022 Apr 29. URL: https://www.census.gov/data [accessed 2025-02-28]
25. All of us research program. All of Us Research Program | NIH. 2020 Jun 1. URL: https://allofus.nih.gov/future-health-begins-all-us [accessed 2024-11-25]
26. Diversity and inclusion. All of Us Research Program | NIH. 2020 Jun 24. URL: https://allofus.nih.gov/about/diversity-and-inclusion [accessed 2025-11-25]
27. Ochieng CA, Minion JT, Turner A, Blell M, Murtagh MJ. What does engagement mean to participants in longitudinal cohort studies? A qualitative study. BMC Med Ethics 2021 Jun 24;22(1):77. [doi: 10.1186/s12910-021-00648-w] [Medline: 34167521]

XSL•FO
RenderX

28. Mapes BM, Foster CS, Kusnoor SV, et al. Diversity and inclusion for the All of Us research program: a scoping review. PLoS ONE 2020;15(7):e0234962. [doi: 10.1371/journal.pone.0234962] [Medline: 32609747]
29. Teague S, Youssef GJ, Macdonald JA, et al. Retention strategies in longitudinal cohort studies: a systematic review and meta-analysis. BMC Med Res Methodol 2018 Nov 26;18(1):151. [doi: 10.1186/s12874-018-0586-7] [Medline: 30477443]
30. Costello L, Dare J, Dontje M, Lambert C, Straker L. Applying the 4Ps of social marketing to retain and engage participants in longitudinal cohort studies: generation 2 Raine study participant perspectives. BMC Med Res Methodol 2022 Nov 5;22(1):288. [doi: 10.1186/s12874-022-01778-4] [Medline: 36335312]
31. Murray AL, Xie T. Engaging adolescents in contemporary longitudinal health research: strategies for promoting participation and retention. J Adolesc Health 2024 Jan;74(1):9-17. [doi: 10.1016/j.jadohealth.2023.06.032] [Medline: 37690009]
32. Ferris EB, Wyka K, Evenson KR, et al. Recruitment and retention strategies for community-based longitudinal studies in diverse urban neighborhoods. JMIR Form Res 2021 Mar 24;5(3):e18591. [doi: 10.2196/18591] [Medline: 33759799]
33. Population studies: diverse cohorts (milestone 1.C). NIH Institute of Aging. URL: https://www.nia.nih.gov/research/milestones/epidemiology-population-studies/population-studies-diverse-cohorts [accessed 2025-02-28]
34. Durocher K, Boparai N, Jankowicz D, Strudwick G. Identifying technology industry-led initiatives to address digital health equity. Digit Health 2021;7:20552076211056156. [doi: 10.1177/20552076211056156] [Medline: 35340562]
35. Participant engagement and health equity workshop. National Institutes of Health (NIH). 2015 Sep 3. URL: https://www.nih.gov/allofus-research-program/participant-engagement-health-equity-workshop [accessed 2024-11-25]
36. Sage bionetworks. Github. URL: https://github.com/Sage-Bionetworks/allofus_engagement [accessed 2025-02-25]

Abbreviations

AoU: All of Us
AoURP: All of Us Research Program
BYOD: bring your own device
EHR: electronic health record
FMH: Family Medical History
HCA: Health Care Access
IRB: institutional review board
LMM: linear mixed model
PMH: Personal Medical History
UBR: underrepresented in biomedical research

Edited by TDA Cardoso; submitted 26.01.24; peer-reviewed by A Hassan, F Albalwy, V Nagesh; revised version received 02.01.25; accepted 17.01.25; published 19.03.25.

Please cite as:
Yadav V, Neto EC, Doerr M, Pratap A, Omberg L
Long-Term Engagement of Diverse Study Cohorts in Decentralized Research: Longitudinal Analysis of "All of Us" Research Program Data
Interact J Med Res 2025;14:e56803
URL: https://www.i-jmr.org/2025/1/e56803
doi:10.2196/56803

© Vijay Yadav, Elias Chaibub Neto, Megan Doerr, Abhishek Pratap, Larsson Omberg. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 19.3.2025. 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

Dengue Epidemiology in 7 Southeast Asian Countries: 24-Year, Retrospective, Multicountry Ecological Study

Shun-Long Weng[1,2*], PhD; Fang-Yu Hung[1*], MD; Sung-Tse Li[2,3*], MD; Bo-Huang Liou[2,4], MD; Chun-Yan Yeung[2,3], PhD; Yu-Lin Tai[3], MD; Yi-Hsuan Wu[4], MD; Ya-Ning Huang[3], MD; Nan-Chang Chiu[2,5], MD; Liang-Yen Lin[6]; Hsin Chi[2,5], PhD; Chien-Yu Lin[2,3], MD

[1]Department of Obstetrics and Gynecology, Hsinchu Municipal MacKay Children's Hospital, Hsinchu, Taiwan

[2]Department of Medicine, MacKay Medical College, New Taipei, Taiwan

[3]Department of Pediatrics, Hsinchu Municipal MacKay Children's Hospital, 28 Jiangong 2nd Rd, East District, Hsinchu, Taiwan

[4]Department of Infectious Disease, Hsinchu MacKay Memorial Hospital, Hsinchu, Taiwan

[5]Department of Pediatric Infectious Disease, MacKay Children's Hospital, Taipei, Taiwan

[6]Department of General Studies, National Hsinchu Girls' Senior High School, Hsinchu, Taiwan

[*]these authors contributed equally

Corresponding Author:
Chien-Yu Lin, MD
Department of Medicine, MacKay Medical College, New Taipei, Taiwan

Abstract

Background: Dengue fever remains the most significant vector-borne disease in Southeast Asia, imposing a substantial burden on public health systems. Global warming and increased international mobility may exacerbate the disease's prevalence. Furthermore, the unprecedented COVID-19 pandemic may have influenced the epidemiological patterns of dengue.

Objective: This study aimed to evaluate epidemiological changes in dengue incidence in Southeast Asia.

Methods: We conducted a retrospective, multicountry ecological study analyzing trends in dengue incidence in 7 Southeast Asian countries from January 2000 to December 2023. Data were extracted from official World Health Organization reports and national health department databases. Countries with data that were incomplete, inconsistent, or not publicly available were excluded from the final analysis. Annual incidence rates were analyzed, and linear trends were calculated to assess long-term patterns.

Results: Epidemiological data from 7 Southeast Asian countries, comprising Thailand, Singapore, Vietnam, Malaysia, the Philippines, Cambodia, and Taiwan, were analyzed across the 24-year study period. A notable nadir in dengue cases was observed coinciding with the COVID-19 pandemic. Significant increasing trends in dengue incidence were identified in Singapore, Vietnam, Malaysia, and the Philippines (slopes: 8.243, 6.513, 8.737, and 8.172; R^2 values: 0.14, 0.34, 0.345, and 0.46, respectively, all $P<.05$).

Conclusions: Dengue fever continues to pose a significant public health challenge in Southeast Asia. Our analysis demonstrates a substantial increase in dengue cases in several countries over the study period. While a temporary decline was observed during the COVID-19 pandemic, a subsequent resurgence of cases highlights the persistent threat of dengue in the region. These findings underscore the critical need for sustained surveillance and innovative control strategies to mitigate the impact of dengue in Southeast Asia.

(Interact J Med Res 2025;14:e70491) doi:10.2196/70491

KEYWORDS

dengue; COVID-19; epidemiology; Southeast Asia; public health

Introduction

Dengue fever, a vector-borne disease transmitted primarily by *Aedes aegypti* and *Aedes albopictus* mosquitoes, remains a significant global health challenge, particularly in Southeast Asia [1]. This arboviral infection presents a wide spectrum of clinical manifestations, ranging from self-limiting febrile illness to severe, life-threatening complications [2]. The global burden of dengue is substantial, with an estimated 400 million infections annually, of which approximately 100 million result in symptomatic disease and 21,000 in fatalities [3,4]. Dengue transmission is closely associated with ambient temperature, and Southeast Asia remains a hyperendemic region [5]. According to a global systematic review of dengue disease

XSL•FO
RenderX

burden, Southeast Asia exhibited significantly higher incidence rates than other world regions [6]. During the decade of 2001 to 2010, the region reported an estimated annual average of 2.9 million dengue cases and 5906 associated deaths [7]. The economic burden, including direct medical costs and productivity loss, along with substantial disability-adjusted life years, further underscores the considerable public health impact in this region. Despite this impact, specific antiviral treatments remain elusive, and effective vaccination strategies are not yet widely implemented, making dengue a persistent threat [8].

The dynamics of dengue transmission are intricately linked to environmental factors, making the disease particularly susceptible to the effects of global climate change. According to the World Meteorological Organization, global mean temperatures have shown a consistent upward trend over the past 2 decades, with an average increase of 0.2 °C per decade [9,10]. This warming trend has significant implications for the epidemiology of vector-borne diseases, including dengue [11,12]. Multiple studies have documented an increase in dengue prevalence correlated with rising temperatures and altered precipitation patterns, suggesting a potential exacerbation of disease burden in the context of ongoing climate change [11-16].

Concurrently, the unprecedented COVID-19 pandemic has dramatically reshaped the landscape of global health, influencing the epidemiological patterns of various infectious diseases [17]. The implementation of stringent public health measures, including lockdowns, international travel restrictions, and nonpharmaceutical interventions, led to a marked reduction in the incidence of several droplet- and contact-transmitted diseases, such as influenza and enterovirus infections. However, the easing of these restrictions has been associated with a resurgence of many infectious diseases, highlighting the complex interplay between public health interventions and disease dynamics [18].

In this context, the epidemiological trends of dengue during and after the COVID-19 pandemic remain largely uncertain. The potential impact of climate change and the disruptive effects of pandemic-related interventions create a unique scenario that warrants thorough investigation. Understanding these trends is crucial for public health planning, resource allocation, and the development of targeted prevention strategies. To address this knowledge gap, we conducted a comprehensive retrospective study examining epidemiological changes in dengue across several Southeast Asian countries from January 2000 to December 2023. The findings of this research have the potential to inform evidence-based policies for dengue control and prevention, contribute to our understanding of the complex interactions between dynamics, pandemics, and vector-borne diseases, and ultimately improve public health outcomes in Southeast Asia and beyond.

Methods

Study Design and Data Sources

We conducted a retrospective, multicountry ecological study analyzing trends in dengue incidence in 7 Southeast Asian countries from January 2000 to December 2023. Data on annual dengue cases were extracted from the following sources: (1) the World Health Organization (WHO) Global Health Observatory data repository [19], (2) the national centers for disease control and prevention of the respective countries, and (3) the ministry of health databases of the selected countries [19-26]. We first screened the WHO Global Health Observatory database to identify countries with available annual dengue incidence data [27]. Subsequently, we cross-validated the data with official national sources, including the websites of ministries of health or national centers for disease control. Countries with data that were incomplete, inconsistent, or not publicly available were excluded from the final analysis. In parallel, we retrieved national annual average temperature trends from publicly accessible climate databases [28]. A flowchart detailing the country selection process is provided in Multimedia Appendix 1.

Data Collection and Processing

For each country, we compiled the annual number of reported dengue cases and population estimates to calculate incidence rates. Crude dengue incidence rates (cases per 100,000 population) were calculated by dividing the annual number of reported cases by the corresponding national population estimates, allowing for cross-country comparability. Population data were obtained from the publicly available United Nations demographic database [29]. Further standardization by age and sex was not performed, as the primary objective of this study was to assess overall national dengue incidence. In cases of data discrepancies between sources, national health authority reports were prioritized, followed by WHO data. Cross-verification between multiple sources was conducted for validation and quality control by 2 authors independently. Dengue is a notifiable communicable disease in all countries included in this study. We used Taiwan as an example, where the diagnosis of dengue is based on a combination of clinical, epidemiological, and laboratory criteria. Genetic typing is performed in specific scenarios, such as the identification of index cases in endemic regions or during outbreak investigations. Confirmed cases are further classified as either imported or locally acquired. In this study, both imported and local cases in Taiwan were included to characterize the composition of reported dengue cases. In addition, national annual average temperature data during the study period were retrieved from publicly accessible meteorological databases.

Statistical Analysis

Statistical analyses were performed using R statistical software (version 4.1.0). Annual incidence rates per 100,000 population were calculated for each country, and epidemiological trends were visualized using the *ggplot2* package. Linear regression models were applied to assess the trends in dengue incidence over time, with the model specified as $Y=\beta0+\beta1X+$, where Y is the annual dengue incidence, X is the year, β0 is the intercept, β1 is the slope, and is the error term. Model parameters, including the slope and R^2 values, were evaluated to determine the direction and strength of the trends. A *P* value less than .05 was considered statistically significant.

XSL•FO
RenderX

Ethical Considerations

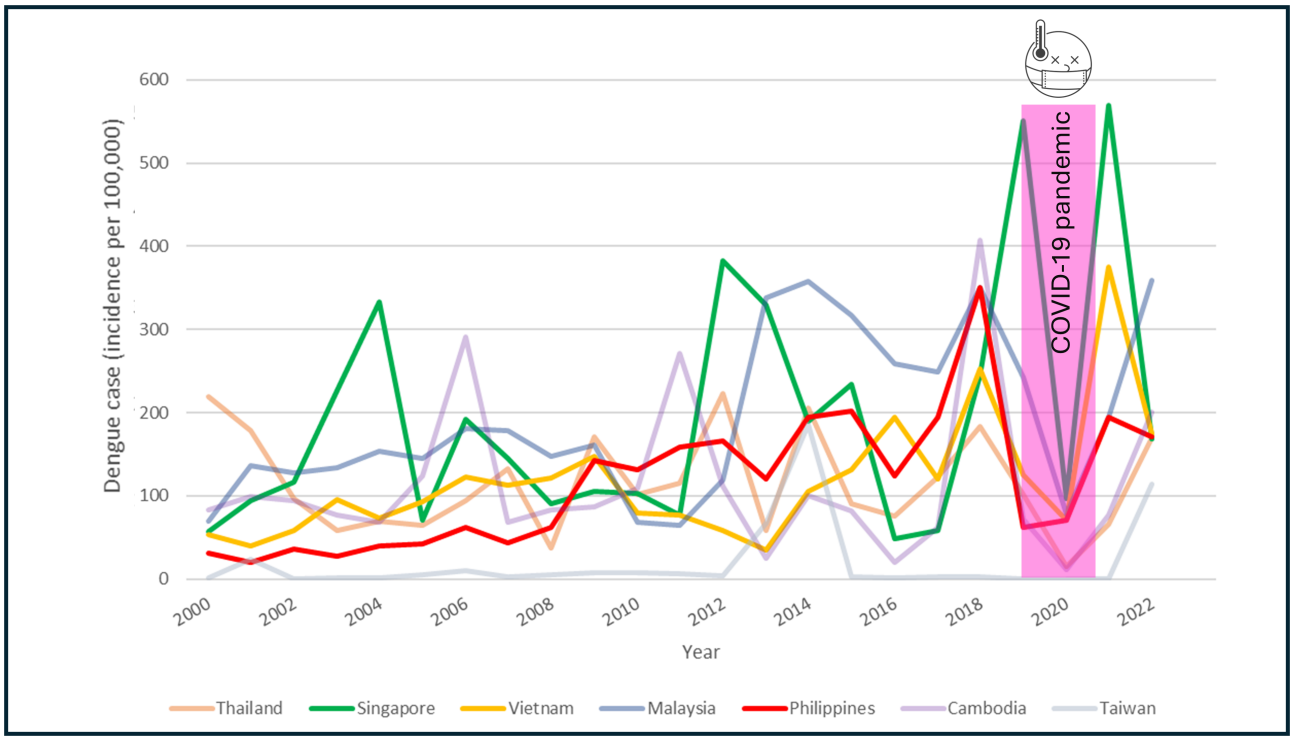
This study was conducted in accordance with the ethical standards of the Declaration of Helsinki [30]. Our study was approved by the institutional review board of MacKay Memorial Hospital, Taipei, Taiwan (20MMHIS140e). This study used publicly available, aggregated data and did not involve individual patient information. We adhered to the principles of ethical research practice throughout the study.

Results

After excluding countries with incomplete data, including Laos, Myanmar, Brunei, Indonesia, and Timor-Leste, 7 countries were included in the final analysis: Thailand, Singapore, Vietnam, Malaysia, the Philippines, Cambodia, and Taiwan. A schematic illustration of the included countries is shown in Multimedia Appendix 2. The annual trends in average temperature across these countries over the 24-year study period are presented in Multimedia Appendix 3, demonstrating an overall temperature increase ranging from 0.97 °C to 2.22 °C. The dynamic trends in dengue cases for the countries are plotted in Figure 1 and Multimedia Appendix 4. Fluctuations in dengue incidence were observed across countries, and dengue was more prevalent in Malaysia, Singapore, and Vietnam. A nadir of dengue cases occurred during the COVID-19 pandemic in 2020, and a resurgence of dengue cases was reported after 2022. Dengue incidence in individual countries is plotted in Figure 2; most countries had a rising incidence during the 24 years. There was no multicountry cluster. The peak incidence was approximately 500 cases per 100,000 inhabitants, observed in Singapore. We further examined the linear trend of dengue incidence in each country and summarize the slopes and R^2 values in Multimedia Appendix 5. All slopes were positive, except for Thailand. Significant increasing trends in dengue incidence were identified in Singapore, Vietnam, Malaysia, and the Philippines (slopes: 8.243, 6.513, 8.737, and 8.172; R^2 values: 0.14, 0.34, 0.35, and 0.46, respectively; all P values <.05).

The epidemiological trend for imported and local cases in Taiwan is plotted in Figure 3. Sporadic imported cases with an increasing trend were observed before COVID-19 (2000 - 2019: slope=0.15; R^2=0.46). A significant surge of local cases was observed in 2014, 2015, and 2023. However, a substantial decline in both imported and local cases was found during the COVID-19 pandemic. A gradual increase in imported cases occurred after the ease of restrictions in 2022, followed by an exponential surge in local cases in 2023.

Figure 1. Dynamic trends in dengue incidence in Southeast Asia.



XSL•FO
RenderX

Figure 2. Trends in dengue incidence in individual countries in Southeast Asia. The inset figure illustrates the annual trends of each country and indicates that no multicountry clusters were observed.

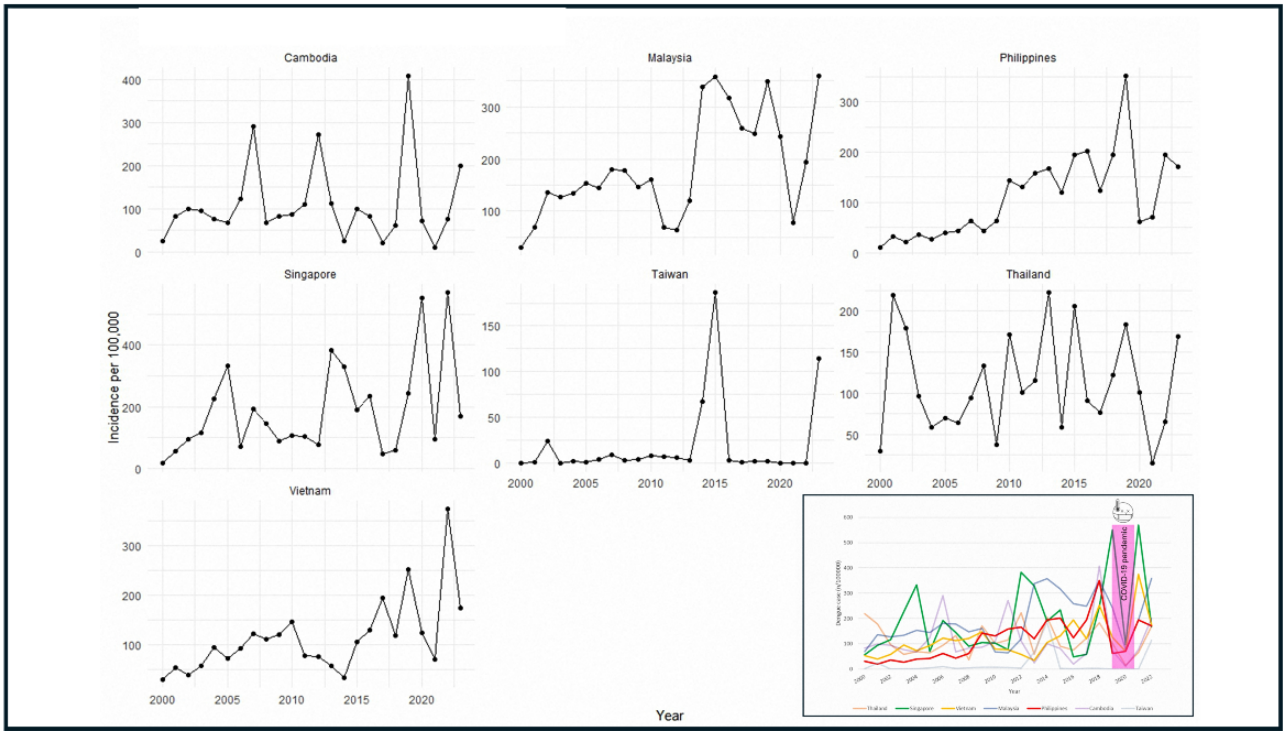
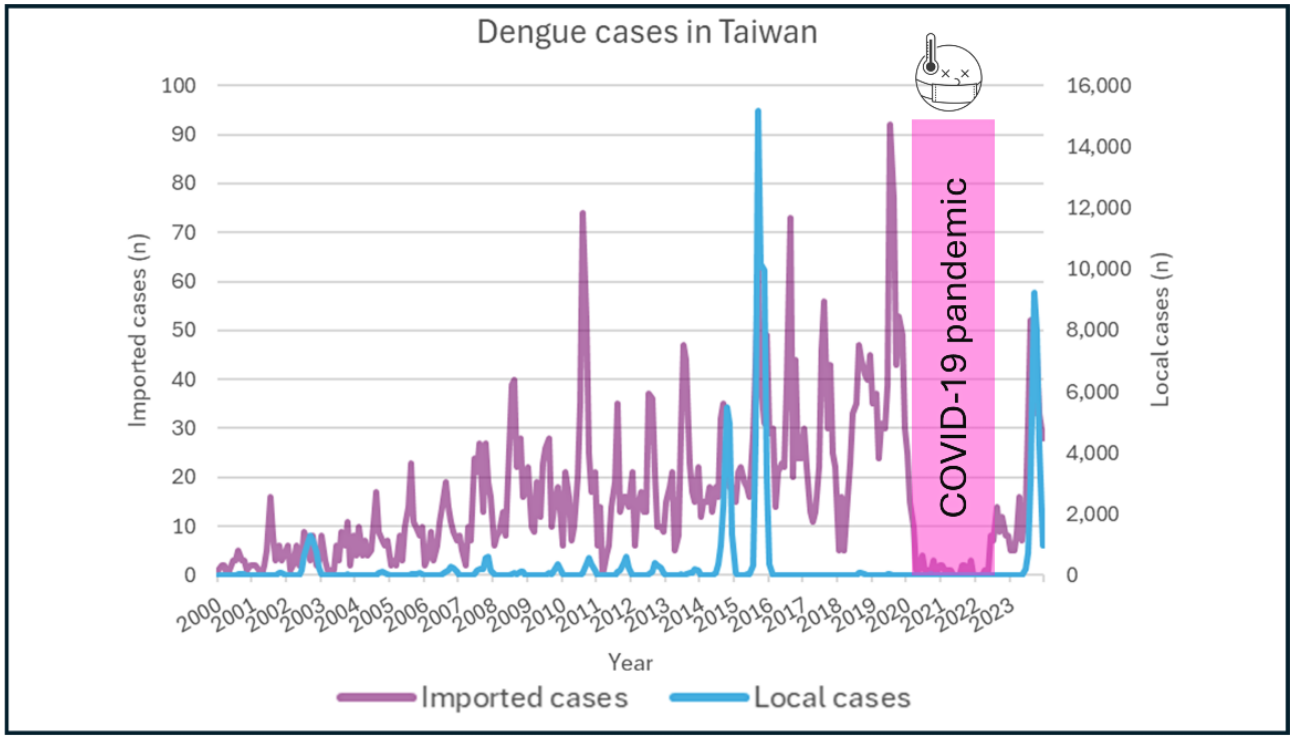


Figure 3. Epidemiological changes in imported and local cases in Taiwan.



Discussion

Principal Findings

The disease burden of dengue is unquestionably substantial. Our study elucidates the epidemiological changes in an endemic region over a 24-year period. Despite advances in public health, we observed a significant increase in dengue incidence in several Southeast Asian countries, including the Philippines, Vietnam, Malaysia, and Singapore. Other countries demonstrated relatively stable endemic patterns. While dengue cases declined during the COVID-19 pandemic, a notable resurgence was observed across all studied countries, underscoring dengue's continued significance as a critical public health threat in the postpandemic era.

A marked nadir in dengue incidence was observed during the COVID-19 pandemic. Public health strategies, implementation

XSL•FO
RenderX

of nonpharmaceutical interventions, and social and international restrictions likely contributed to this observed reduction. In Sri Lanka, for instance, national dengue case numbers significantly decreased during lockdown periods [31]. In Australia, dengue cases nearly disappeared during the pandemic [32]. Our study revealed that Thailand, Cambodia, and Taiwan reported fewer than 30 cases per 100,000 inhabitants during this period. However, the apparent decrease in reported dengue cases may be attributable to underreporting due to several factors: inadequate medical resources, compromised communication systems, limited diagnostic capabilities, and reduced health care–seeking behavior. Most dengue infections present with mild symptoms such as fever, headache, arthralgia, and skin rashes, and individuals with mild illness may have been less likely to seek medical attention due to reduced health care services and concerns about potential COVID-19 exposure. Additionally, health care providers were primarily focused on COVID-19 diagnostics, potentially leading to decreased dengue screening and testing. During the COVID-19 pandemic, neither the virus nor vector-borne mosquitoes disappeared; instead, they remained latent within communities, waiting for opportunities to resurface. In the aftermath of the COVID-19 pandemic, an "immune debt" phenomenon has emerged, leading to the resurgence of numerous infectious diseases, including influenza, enterovirus, respiratory syncytial virus, and dengue fever. Multiple countries, such as Brazil, Australia, Egypt, and Taiwan, have experienced significant dengue fever outbreaks following the relaxation of pandemic-related restrictions [32-37]. Our research has uncovered similar trends: during the COVID-19 pandemic, stringent preventive measures and reduced international travel disrupted typical epidemiological patterns. Consequently, the health threat posed by dengue fever remains a critical public health concern in the postpandemic era.

Climate change and global warming may contribute to the expanded distribution of vector mosquitoes and increased dengue incidence [11,12,35,38-41]. The optimal temperature for *Aedes albopictus* and *Aedes aegypti* is approximately 25 °C to 30 °C [10]. Warmer temperatures can shorten viral replication incubation periods in mosquitoes, accelerate mosquito growth, and increase biting frequency. Our study showed an overall temperature increase ranging from 0.97 °C to 2.22 °C. Consequently, climate change may expand vector habitat ranges and potentially increase infection transmission zones. Theoretically, dengue is more prevalent in tropical regions. Our study showed a relatively lower incidence in Taiwan, a subtropical country. However, dengue prevalence is influenced by multiple complex factors. We found that the dengue incidence rates significantly increased over the study period in the Philippines, Vietnam, Malaysia, and Singapore. However, these countries were not those with the lowest latitudes or the highest average temperatures among the study sample. Urbanization may reduce breeding sites, alter water accumulation patterns, and impact vector proliferation. Population density and international travel resumption potentially contribute to disease spread [42]. Interestingly, some research has explored correlations between economic indicators like gross domestic product (GDP) and dengue transmission, though findings remain inconsistent [43,44]. Countries with lower GDP often have limited investment in public health infrastructure, medical services, and vector control programs, which may contribute to higher dengue transmission. Conversely, widespread dengue outbreaks may negatively impact national economies by reducing tourism revenue and workforce productivity, thereby further decreasing GDP. In brief, the incidence of dengue epidemics is fueled by multiple complex factors [45,46]. Data from the past 24 years indicate that these evolving trends cannot be explained by any single factor alone.

Limitations

Our study possesses notable strengths, primarily stemming from its comprehensive analysis of national databases that systematically tracked epidemiological trends across 7 countries over an extensive 24-year period, encompassing both pre- and post–COVID-19 pandemic epochs. However, the research is not without limitations. The study was unable to comprehensively characterize the severity of dengue infections, which range from asymptomatic to severe. Moreover, multiple complex and potentially influential factors remained unaccounted for in this analysis, including nuanced variations in climate patterns, urbanization levels, international travel dynamics, economic indicators, and vaccination status. These unexamined variables could potentially modulate dengue transmission in ways not fully captured by our current methodological approach. Due to the unavailability of reliable data, several countries were excluded from the final analysis. Consequently, while our findings provide valuable insights into regional dengue epidemiology, they also underscore the necessity for future research that can more comprehensively integrate these multifaceted determinants of disease transmission.

Conclusion

Our 24-year epidemiological analysis of dengue fever in Southeast Asia reveals the complex dynamics of disease transmission. The COVID-19 pandemic temporarily disrupted transmission patterns, but the subsequent resurgence highlights the virus's resilience. Ongoing, multicountry surveillance efforts are essential for the effective prevention and control of dengue.

Acknowledgments

We sincerely thank the World Health Organization and the respective national health authorities for their efforts in disease surveillance and data collection, which made this research possible. We acknowledge the use of generative language tools (ChatGPT and Claude) for language translation and polishing.

XSL•FO
RenderX

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SLW, FYH, and CYL were responsible for conception and draft writing. STL, BHL, CYY, YLT, LYL, and HC performed the critical review and literature search. STL, YHW, YNH, and NCC were responsible for statistics, supervision, and validation. All authors were involved in data collection and analysis and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Flowchart of database search and enrollment.
[PNG File, 640 KB - i-jmr_v14i1e70491_app1.PNG]

Multimedia Appendix 2
Geographic illustration of enrolled countries.
[PNG File, 1166 KB - i-jmr_v14i1e70491_app2.PNG]

Multimedia Appendix 3
Average temperature trends during the study period.
[PNG File, 628 KB - i-jmr_v14i1e70491_app3.PNG]

Multimedia Appendix 4
Dengue cases and incidences per 100,000 inhabitants in Southeast Asia (2000-2023).
[DOCX File, 26 KB - i-jmr_v14i1e70491_app4.docx]

Multimedia Appendix 5
The linear trend, slopes, and R^2 values of dengue incidence in each country between 2000 and 2023.
[DOCX File, 17 KB - i-jmr_v14i1e70491_app5.docx]

References

1. Paz-Bailey G, Adams LE, Deen J, Anderson KB, Katzelnick LC. Dengue. Lancet 2024 Feb 17;403(10427):667-682. [doi: 10.1016/S0140-6736(23)02576-X] [Medline: 38280388]
2. Leandro A, Maciel-de-Freitas R. Development of an integrated surveillance system to improve preparedness for arbovirus outbreaks in a dengue endemic setting: descriptive study. JMIR Public Health Surveill 2024 Nov 14;10:e62759. [doi: 10.2196/62759] [Medline: 39588736]
3. Sarker R, Roknuzzaman ASM, Haque MA, Islam MR, Kabir ER. Upsurge of dengue outbreaks in several WHO regions: public awareness, vector control activities, and international collaborations are key to prevent spread. Health Sci Rep 2024 Apr;7(4):e2034. [doi: 10.1002/hsr2.2034] [Medline: 38655420]
4. Simmons CP, Farrar JJ, Nguyen VV, Wills B. Dengue. N Engl J Med 2012 Apr 12;366(15):1423-1432. [doi: 10.1056/NEJMra1110265] [Medline: 22494122]
5. Deng J, Zhang H, Wang Y, et al. Global, regional, and national burden of dengue infection in children and adolescents: an analysis of the Global Burden of Disease Study 2021. EClinicalMedicine 2024 Dec;78:102943. [doi: 10.1016/j.eclinm.2024.102943] [Medline: 39640938]
6. Ahmad Zamzuri MAI, Md Hanif S, Halim A, Rafi'i M, Asari S, Hod R, et al. Systematic review of the economic burden of dengue infection to the healthcare in South East Asia (SEA). Int J Public Health Res 2025 Feb;15(1):2087-2104 [FREE Full text]
7. Shepard DS, Undurraga EA, Halasa YA. Economic and disease burden of dengue in Southeast Asia. PLOS Negl Trop Dis 2013;7(2):e2055. [doi: 10.1371/journal.pntd.0002055] [Medline: 23437406]
8. Yang X, Quam MBM, Zhang T, Sang S. Global burden for dengue and the evolving pattern in the past 30 years. J Travel Med 2021 Dec 29;28(8):taab146. [doi: 10.1093/jtm/taab146] [Medline: 34510205]
9. Taiwan Central Weather Administration. Climate data. URL: https://www.cwa.gov.tw/eng/ [accessed 2025-08-22]
10. World Meteorological Organization. URL: https://wmo.int [accessed 2025-08-22]

XSL•FO
RenderX

11. Bellone R, Lechat P, Mousson L, et al. Climate change and vector-borne diseases: a multi-omics approach of temperature-induced changes in the mosquito. J Travel Med 2023 Jun 23;30(4):taad062. [doi: 10.1093/jtm/taad062] [Medline: 37171132]
12. Edelson PJ, Harold R, Ackelsberg J, et al. Climate change and the epidemiology of infectious diseases in the United States. Clin Infect Dis 2023 Mar 4;76(5):950-956. [doi: 10.1093/cid/ciac697] [Medline: 36048507]
13. Cheong YL, Mohd Ghazali S, Mat Hashim MH, et al. Exploring 97 years of Aedes aegypti as the Vector for dengue, yellow fever, zika, and chikungunya (Diptera: Culicidae): scientometric analysis. Interact J Med Res 2025 Apr 23;14:e65844. [doi: 10.2196/65844] [Medline: 40267478]
14. Childs ML, Lyberger K, Harris M, Burke M, Mordecai EA. Climate warming is expanding dengue burden in the americas and asia. Infectious Diseases (except HIV/AIDS). Preprint posted online on Jan 9, 2024. [doi: 10.1101/2024.01.08.24301015]
15. Chiu NC, Chi H, Weng SL, Lin CY. Unprecedented dengue outbreak in Taiwan following COVID-19. J Travel Med 2024 Mar 1;31(2):taae015. [doi: 10.1093/jtm/taae015] [Medline: 38246750]
16. Ritu MR, Sikder D, Patwary MM, Tamim AR, Rodriguez-Morales AJ. Climate change, urbanization and resurgence of dengue in Bangladesh. New Microbes New Infect 2024 Jun;59(101414):101414. [doi: 10.1016/j.nmni.2024.101414] [Medline: 38660578]
17. Chiu NC, Chi H, Tai YL, et al. Impact of wearing masks, hand hygiene, and social distancing on influenza, enterovirus, and all-cause pneumonia during the coronavirus pandemic: retrospective national epidemiological surveillance study. J Med Internet Res 2020 Aug 20;22(8):e21257. [doi: 10.2196/21257] [Medline: 32750008]
18. Chiu NC, Huang DTN, Weng SL, et al. Resurgence of influenza and enterovirus infections in Taiwan post-COVID-19: a nationwide surveillance study. J Infect Public Health 2024 Nov;17(11):102560. [doi: 10.1016/j.jiph.2024.102560] [Medline: 39418957]
19. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/En [accessed 2025-08-22]
20. Dengue. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue [accessed 2025-08-22]
21. Department of Health, Ministry of Public Health, Thailand. URL: https://moph.go.th [accessed 2025-08-25]
22. Republic of the Philippines Department of Health. URL: https://doh.gov.ph [accessed 2025-08-22]
23. Kementerian Kesihatan Malaysia. URL: https://www.moh.gov.my [accessed 2025-08-22]
24. Ministry of Health, Cambodia. URL: https://moh.gov.kh/kh/home [accessed 2025-08-22]
25. Ministry of Health, Singapore. URL: https://www.moh.gov.sg [accessed 2025-08-22]
26. Ministry of Health, Vietnam. URL: https://moh.gov.vn/web/ministry-of-health [accessed 2025-08-22]
27. Dengue: latest situation reports. World Health Organization. URL: https://www.who.int/westernpacific/wpro-emergencies/surveillance/dengue [accessed 2025-08-22]
28. Data overview. Berkeley Earth. URL: https://berkeleyearth.org/data [accessed 2025-08-22]
29. World population prospects. UN Department of Economic and Social Affairs. URL: https://population.un.org/wpp [accessed 2025-08-22]
30. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
31. Liyanage P, Rocklöv J, Tissera HA. The impact of COVID-19 lockdown on dengue transmission in Sri Lanka; a natural experiment for understanding the influence of human mobility. PLoS Negl Trop Dis 2021 Jun;15(6):e0009420. [doi: 10.1371/journal.pntd.0009420] [Medline: 34111117]
32. Sohail A, Anders KL, McGuinness SL, Leder K. The epidemiology of imported and locally acquired dengue in Australia, 2012-2022. J Travel Med 2024 Mar 1;31(2):taae014. [doi: 10.1093/jtm/taae014] [Medline: 38243558]
33. Naveca FG, Santiago GA, Maito RM, et al. Reemergence of dengue virus serotype 3, Brazil, 2023. Emerg Infect Dis 2023 Jul;29(7):1482-1484. [doi: 10.3201/eid2907.230595] [Medline: 37347884]
34. Pendrey CG, Strachan J, Peck H, et al. The re-emergence of influenza following the COVID-19 pandemic in Victoria, Australia, 2021 to 2022. Euro Surveill 2023 Sep;28(37):2300118. [doi: 10.2807/1560-7917.ES.2023.28.37.2300118] [Medline: 37707981]
35. Taylor L. Dengue fever: Brazil rushes out vaccine as climate change fuels unprecedented surge. BMJ 2024;38408778:q483. [doi: 10.1136/bmj.q483] [Medline: 38408778]
36. García-García D, Fernández-Martínez B, Bartumeus F, Gómez-Barroso D. Modeling the regional distribution of international travelers in Spain to estimate imported cases of dengue and malaria: statistical inference and validation study. JMIR Public Health Surveill 2024 May 27;10:e51191. [doi: 10.2196/51191] [Medline: 38801767]
37. Manciulli T, Zammarchi L, Lagi F, et al. Emergence of dengue fever: sentinel travellers uncover outbreak in Sharm El-Sheikh, Egypt, May 2024. J Travel Med 2024 Jul 7;31(5):taae080. [doi: 10.1093/jtm/taae080] [Medline: 38857380]
38. Araf Y, Akter M, Zhai J, Zheng C, Hossain MG. Emerging health implications of climate change: dengue outbreaks and beyond in Bangladesh. Lancet Microbe 2024 Mar;5(3):e213. [doi: 10.1016/S2666-5247(23)00366-X] [Medline: 38219754]
39. Barcellos C, Matos V, Lana RM, Lowe R. Climate change, thermal anomalies, and the recent progression of dengue in Brazil. Sci Rep 2024 Mar 11;14(1):5948. [doi: 10.1038/s41598-024-56044-y] [Medline: 38467690]

XSL•FO
RenderX

40. Islam MT, Kamal A, Islam MM, Hossain S. Impact of climate change on dengue incidence in Singapore: time-series seasonal analysis. Int J Environ Health Res 2024 Dec;34(12):3988-3998. [doi: 10.1080/09603123.2024.2337827] [Medline: 38627938]
41. Wang Y, Li C, Zhao S, et al. Projection of dengue fever transmissibility under climate change in South and Southeast Asian countries. PLoS Negl Trop Dis 2024 Apr;18(4):e0012158. [doi: 10.1371/journal.pntd.0012158] [Medline: 38683870]
42. Ganser I, Thiébaut R, Buckeridge DL. Global variations in event-based surveillance for disease outbreak detection: time series analysis. JMIR Public Health Surveill 2022 Oct 31;8(10):e36211. [doi: 10.2196/36211] [Medline: 36315218]
43. Harish V, Colón-González FJ, Moreira FRR, et al. Human movement and environmental barriers shape the emergence of dengue. Nat Commun 2024 May 28;15(1):4205. [doi: 10.1038/s41467-024-48465-0] [Medline: 38806460]
44. Lee SY, Shih HI, Lo WC, Lu TH, Chien YW. Discrepancies in dengue burden estimates: a comparative analysis of reported cases and Global Burden of Disease Study, 2010-2019. J Travel Med 2024 Jun 3;31(4):taae069. [doi: 10.1093/jtm/taae069] [Medline: 38696416]
45. Sylvestre E, Cécilia-Joseph E, Bouzillé G, et al. The role of heterogenous real-world data for dengue surveillance in Martinique: observational retrospective study. JMIR Public Health Surveill 2022 Dec 22;8(12):e37122. [doi: 10.2196/37122] [Medline: 36548023]
46. Colón-González FJ, Gibb R, Khan K, Watts A, Lowe R, Brady OJ. Projecting the future incidence and burden of dengue in Southeast Asia. Nat Commun 2023 Sep 6;14(1):5439. [doi: 10.1038/s41467-023-41017-y] [Medline: 37673859]

Abbreviations

GDP: gross domestic product
WHO: World Health Organization

Edited by TDA Cardoso; submitted 24.12.24; peer-reviewed by J Nejati, N Jain, SMA Hasan; revised version received 26.07.25; accepted 29.07.25; published 08.09.25.

Please cite as:
Weng SL, Hung FY, Li ST, Liou BH, Yeung CY, Tai YL, Wu YH, Huang YN, Chiu NC, Lin LY, Chi H, Lin CY
Dengue Epidemiology in 7 Southeast Asian Countries: 24-Year, Retrospective, Multicountry Ecological Study
Interact J Med Res 2025;14:e70491
URL: https://www.i-jmr.org/2025/1/e70491
doi:10.2196/70491

© Shun-Long Weng, Fang-Yu Hung, Sung-Tse Li, Bo-Huang Liou, Chun-Yan Yeung, Yu-Lin Tai, Yi-Hsuan Wu, Ya-Ning Huang, Nan-Chang Chiu, Liang-Yen Lin, Hsin Chi, Chien-Yu Lin. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 8.9.2025. 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

Factors Predicting Information Overload During the COVID-19 Pandemic in the Digital Age: Longitudinal Study

Hiroko Okada[1]; Tsuyoshi Okuhara[1]; Takahiro Kiuchi[1]

Department of Health Communication, Schoolf of Public Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

Corresponding Author:
Hiroko Okada
Department of Health Communication, Schoolf of Public Health
Graduate School of Medicine
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo, 113-8655
Japan
Phone: 81 3 5800 6549
Email: okadahiroko-tky@umin.ac.jp

Abstract

Background: The human capacity to process information is limited. During the COVID-19 pandemic, people were exposed to a large amount of uncertain and complex health information. This situation made some people experience perceived information overload, which made them unable to adopt appropriate preventive behaviors.

Objective: This study aimed to examine the individual characteristics, abilities, and attention to informational media that predict the perception of information overload during a pandemic.

Methods: We conducted a longitudinal study with 2 time points, August 2020 and August 2021, among residents of Japan under a COVID-19 emergency declaration. The sample had the same proportions for sex, age, and prefecture as the general Japanese population. We used a web-based survey to measure sociodemographic characteristics, health literacy (HL), attention to 6 different types of information channels, and participants' perception of information overload. Hierarchical multiple regression analysis was conducted with information overload as the objective variable.

Results: A total of 784 participants responded to the survey at both time points, with a follow-up rate of 78.4% (784/1000). Hierarchical multiple regression analysis showed that younger age (β=−0.084, 95% CI −0.142 to −0.013), male sex (β=−0.163, 95% CI −0.008 to −0.003), lower HL (β=−0.084, 95% CI −0.114 to −0.011), paying less attention to television news (β=−0.118, 95% CI −0.038 to −0.001), and paying greater attention to social media (β=0.089, 95% CI 0.000-0.027) significantly predicted information overload 1 year after exposure to information during the pandemic.

Conclusions: Public health communicators should aim to provide concise and understandable information in consideration of a target population that is vulnerable to information overload during a pandemic. A high level of attention to social media may increase the perception of information overload. By contrast, HL may reduce the cognitive load in information processing. Providing an environment during normal periods that allows people to develop the skills to critically interpret health information will help them to prepare for future infodemics.

(Interact J Med Res 2025;14:e67098) doi:10.2196/67098

KEYWORDS

COVID-19; health communication; information overload; social media; infobesity; infoxication; information anxiety; data overload; SARS-CoV-2; coronavirus; respiratory; infectious; pulmonary; pandemic; digital age; information age; longitudinal study; Japan; human capacity; informational media

XSL•FO
RenderX

Introduction

Background

The unexpected occurrence of the COVID-19 pandemic caused a serious and prolonged threat to human health and life. To cope with this threat, people consumed information from diverse media sources to stay informed about this emerging infectious disease [1,2]. At the 2020 Munich Security Conference, the World Health Organization (WHO) director general stated that the world was not only dealing with a pandemic, but also with an infodemic [3,4]. The COVID-19 pandemic occurred amid the ongoing proliferation of new media platforms, such as social media, where anyone can disseminate any type of information. Fake news, misinformation, and conspiracy theories spread faster than ever via these new media channels [3,5,6]. In addition, information during a pandemic is often uncertain. It has become apparent that ensuring that the public can adequately perceive information to protect their own lives and well-being during a pandemic is an enormous challenge within public health communication.

Information Overload

The human capacity to process information is limited [7-10]. Therefore, a rapid influx of information can impose a psychological burden. Information overload can be defined as a situation where the amount of information provided within a certain time frame exceeds an individual's capacity to process the information [11]. Information overload affects people's decision-making accuracy, time to reach a decision, and general performance [12]. More specifically, with information overload, a person can become overly selective, ignore large amounts of information, have difficulty understanding the relevance of details with respect to the overall perspective, and require more time and effort to acquire information [12]. Similarly, when people are faced with overly complex health information, they may not recognize the information as personally relevant or valuable, and may be unable to infer what information is needed to engage in appropriate behavior [13].

Information Overload and Health Behavior

People experiencing information overload tend to have negative attitudes toward health information [14,15]. Consequently, they are less motivated to seek out and digest such information, which ultimately reduces their health knowledge and associated skills [15,16]. In the field of health, the problem of information overload has been studied in recent years. For example, a study of cancer information seeking showed that information overload leads to the avoidance of cancer screening and a lack of appropriate preventive behavior [17]. A study of patients with newly diagnosed coronary artery disease suggested that information overload has a direct negative impact on their intention to read self-management materials [18]. Previous cross-sectional studies conducted during the COVID-19 pandemic showed that information overload is associated with vaccine hesitancy [19,20]. Information overload in the COVID-19 era has also been reported to cause negative emotions, such as anxiety and depression [21,22]. These findings show that information overload is a health communication concern as a factor that influences health behavior.

Predictors of Information Overload: Theoretical Framework

Cognitive load theory is an important framework for understanding the causes of information overload [23]. This theory was originally introduced in the field of educational psychology and has contributed to the development of educational environments for the retention of knowledge and information [24]. As an application of learning theory, it is also useful for understanding the process by which humans acquire knowledge from news and other sources. In this theory, cognitive load is composed of 3 conceptualized factors: intrinsic, extraneous, and germane load [24]. Intrinsic cognitive load is related to the natural complexity of the information or learning materials that must be understood with respect to a specific topic. Extraneous load can be expressed as the load imposed by information from suboptimal media. When extraneous cognitive load is high, the information seeker disproportionately uses their working memory (WM) resources to deal with the provided media rather than with information seeking itself, so the attention devoted to essential information is reduced. Germane load refers to the WM resources that the information seeker expends to deal with the intrinsic cognitive load of the information. Germane load is related to individual characteristics [24].

Not all individuals experience information overload. The relationship between age and cognitive function, an individual characteristic, has been the focus of research in the fields of neuroscience and gerontology [25-27]. Related studies have shown that cognitive function declines with age in adulthood [28,29]. This suggests that the cognitive load from processing complex information in adulthood is greater with older age. However, a cross-sectional study conducted in Korea during the COVID-19 pandemic found no association between age and information overload [30]. The relationship between age and information overload in the context of health information and pandemics has not been clarified. However, evidence from the field of neuroscience shows that the cognitive load of information increases with age. It is likely that older people were more likely to perceive information overload during the pandemic, when they were exposed to a large amount of complex health information.

Our first hypothesis (hypothesis 1) is that older adults are more likely to experience information overload 1 year later in a pandemic information environment.

Previous studies have indicated that men and women have different psychological adaptations to extreme situations, such as pandemic lockdowns [31]. The WHO has emphasized the differential impact of a pandemic on men and women and the need for research on the effects of the COVID-19 pandemic from a sex-based perspective [32]. The relationship between information overload and sex has been reported in contexts of exposure to news. These studies have shown that women tend to experience greater information overload than men owing to news exposure [4,33,34]. It has been found that men and women use different cognitive resources with respect to news story structures, with women having slower reaction times to news stories [35]. The sex differences in cognitive load are also

XSL•FO
RenderX

supported by research reporting functional magnetic resonance imaging data [26]. When men and women were given the same verbal and visuospatial WM tasks, women experienced greater task demands and higher cognitive load than men. However, as mentioned, the relationship between sex and information overload in the context of health information and pandemics remains unclear. Studies conducted in the context of exposure to early COVID-19 and cancer information reported no significant differences in the perception of information overload between men and women [36]. These studies also had limitations, such as having a cross-sectional design or having limited participant demographics. In the context of the COVID-19 pandemic, news continues to be disseminated daily. Evidence regarding the processing of WM tasks and news suggests that women may be more likely to recognize information overload in the information environment within a pandemic context.

Our first hypothesis (hypothesis 2) is that women are more likely to experience information overload 1 year later in a pandemic information environment.

Health information frequently contains complex medical wording [37,38]. This can especially affect the processing of information, especially for people with reading and writing difficulties [39]. Furthermore, the information released during a pandemic is fraught with uncertainty and is continually revised as new findings emerge. People must engage in repeated information processing each time information is updated. The intrinsic cognitive load caused by this complexity of information is affected by the learner's previous psychological state and knowledge [23]. Therefore, it has been reported that having relevant previous specialist knowledge reduces the learner's intrinsic cognitive load [40]. The degree to which an individual is able to find, understand, and use information and services that are useful for making decisions and taking action regarding their health is referred to as health literacy (HL). There are 3 aspects to HL. Functional HL refers to basic reading and writing skills, communicative HL refers to the ability to extract information from various forms of communication and apply it to changing situations, and critical HL refers to the ability to critically evaluate and use information to exert greater influence [41]. HL has been noted to be related to the experience of information overload as well as health status [41]. For example, previous studies conducted in the context of information seeking regarding cancer and weight management have reported that lower HL correlates with a higher perception of health information overload [42-44]. A cross-sectional study conducted in the United States during the COVID-19 pandemic also reported that individuals with lower HL were more likely to perceive information overload [45]. It is rational to consider that longitudinal analyses would similarly find that high levels of HL are likely to prevent information overload.

Our first hypothesis (hypothesis 3) is that people with higher HL are less likely to experience information overload 1 year later in a pandemic information environment.

When faced with an unknown infectious disease, people pay attention to information provided through various types of media to fill gaps in their knowledge [1,2]. It has been reported that processing information presented by diverse media sources is cognitively difficult [46]. There have been various reports on the impact of social media on users' cognitive load. For example, there are reports on intrinsic cognitive load factors, such as fear of missing out (FOMO), privacy concerns, anxiety, and depression [47]. There have also been reports of extrinsic cognitive load factors, such as parental influence, cyberbullying, complexity, technology-related factors, and social overload [48]. In a study examining news attention and information overload, television viewing lowered the perception of information overload, whereas the use of a PC or Facebook increased perceptions of information overload [33]. During the COVID-19 pandemic, social media has been used daily as a platform for information exchange [49]. Social media users are constantly searching, receiving, identifying, reading, and processing a continuous flow of information. Users have been reported to experience a drain on cognitive energy, feeling overwhelmed and fatigued by the cognitive demands of social media [48,50]. Several cross-sectional studies have reported that people who follow COVID-19-related information on social media are more likely to experience information overload [51,52]. It is rational to assume that a longitudinal analysis would find that attention to social media leads to a greater perception of information overload than exposure to old media.

Our first hypothesis (hypothesis 4) is that greater exposure to social media information increases the experience of information overload 1 year later in a pandemic information environment.

The information flow during a pandemic differs in quantity and quality compared with that during a normal search for information on diseases, such as cancer. There is a need to identify the characteristics of individuals and information channels that predict information overload during a pandemic. Combined with skills that can reduce cognitive load owing to complex information, this information would provide important resources for the design of public health communication strategies to increase appropriate preventive behaviors. However, to the best of our knowledge, no previous studies have longitudinally examined the predictors of information overload during the COVID-19 pandemic.

This Study

The purpose of this study was to longitudinally examine predictors of information overload during the COVID-19 pandemic in terms of individual characteristics, health information–handling skills, and exposure to information as an available resource for information processing. Identifying the factors that influence information overload is essential to the theoretical understanding of these associations and to improving the efficiency of public health communication.

Methods

Study Design

We conducted a longitudinal study in Japan using a web-based survey administered at 2 time points: in August 2020, soon after the WHO's declaration of the COVID-19 pandemic, and 1 year later, in August 2021. The state of emergency declared by the Japanese government was in effect at both times.

XSL•FO
RenderX

Data Collection

A sample was drawn from a panel of 2.2 million Japanese residents using the platform of the research firm Rakuten Insight, Inc. This study was conducted as part of a multipurpose longitudinal survey to examine multiple research questions during the pandemic. We used nonprobabilistic quota sampling for age, gender, and place of residence to match the Japanese population distribution. The sample had a similar distribution for sex, age, and prefecture as the general Japanese population. People eligible to participate in the study were those (1) who lived in areas where a state of emergency had been declared at our study time points, (2) who were not infected with COVID-19 and had no family members who were infected, and (3) who were aged 18 years or older.

We conducted the first survey (T1) on August 15 and 16, 2020, and the second survey (T2) on August 15 to 17, 2021. The 1820 members of the survey panel who met the inclusion criteria were invited via email to respond to the screening questions. As part of the screening process, participants were required to read a web page that outlined our research. After agreeing to participate, participants were given a URL to access the online questionnaire, and then they completed the questionnaire manually. This survey was a closed survey that required participants to enter their ID and password to complete. The survey form developed in this study was sent to 10 participants who had been recruited using the convenience sampling method, and we checked for any items that were difficult to answer or inappropriate before the main survey. Using adaptive questioning, we minimized the number of questions displayed to individuals. All questions displayed were mandatory, and the next question was displayed after the participant had completed the previous question. There was a page where participants could review their answers after completing all the questions, and they could make corrections before submitting their answers.

The T1 closed when 1000 participants from 8 prefectures (Chiba, Hokkaido, Hyogo, Kanagawa, Kyoto, Osaka, Saitama, and Tokyo) completed the questionnaire. The T2 was conducted among the participants recruited in T1. The areas surveyed were under a state of emergency during both survey periods. An analysis was conducted among the 784 participants who completed both surveys. The follow-up rate was 78.4% (784/1000).

Measurements

Independent Variables

Sociodemographic Data

Data on age (continuous), sex (0=male, 1=female), level of education (categorical), household income (categorical), history of infection (binary), and cohabiting family (binary) were collected during T1.

Health Literacy

HL was measured using the 5-item version of the Communicative Critical Health Literacy instrument developed by Ishikawa et al [53]. This instrument uses a 5-point Likert scale ranging from 1 (completely disagree) to 5 (strongly agree) and was applied in T1. The score was the average of 5 items (continuous variables), with higher scores indicating higher HL. The SD in this study was 0.6. The Cronbach α in this study was 0.88.

Attention Paid to Information Channels About the COVID-19 Pandemic

The T1 assessed the frequency of exposure to 7 types of media information channels: newspapers, television news, television tabloid shows, news websites, official websites of the government and medical professional organizations, video-sharing sites, and social media. For each channel, respondents were asked a single question about the extent to which they had paid attention to COVID-19–related information on that media channel over the past week (eg, “How much attention did you pay to information about COVID-19 in newspapers?”). Each item was rated on a 10-point scale that ranged from 1 (paid no attention at all) to 10 (paid a great deal of attention), which was adapted from a previous study [53]. The score was a continuous variable ranging from 1 to 10, with higher scores indicating greater attention to that information channel.

COVID-19 Information Overload

Information overload was assessed using the Corona Information Overload Scale developed by Sarkhel et al [54]. This scale is an adaptation of the Cancer Information Overload scale, modified for the COVID-19 setting [54]. Example items include statements, such as “There are so many different recommendations for the prevention of COVID-19 that it is difficult to know which ones I should follow.” The scale comprises 8 items, with each item rated on a 4-point Likert scale ranging from 1 (do not at all agree) to 4 (very much agree). The score is the average of 8 items (continuous variables), with higher scores indicating a greater perception of information overload. Information overload was assessed at T2. The SD in this study was 0.4. The Cronbach α in this study was 0.89.

Ethical Considerations

A protocol for this study was approved by the ethics review committee of the Graduate School of Medicine, University of Tokyo, based on guidelines for clinical trials involving human subjects (11270). Participants were provided with an explanation of the study and completed a web-based consent form. All participants were given the right to withdraw from the study at any time during the study period. Data were provided to researchers by the research company in a linked but anonymized form. Participants in the online survey were offered points that could be used as 1 yen (US $0.009) per point on the Rakuten e-commerce site as an incentive.

Statistical Analysis

Descriptive statistics were calculated for each sociodemographic variable. We then performed single linear regression analysis as a crude analysis, with the objective variables being information overload, sociodemographic characteristics, HL, and attention to media information.

Hierarchical multiple linear regression analysis was performed as an autoregressive model. In model 1, we included information overload (T2) as the outcome variable and sociodemographic

XSL•FO
RenderX

characteristics, including age (continuous; sex, 0=male, 1=female), level of education (categorical), household income (categorical), history of infection (binary), and cohabiting family (binary), as independent variables. In addition to the variables in model 1, we included HL (T1) as an independent variable in model 2. In addition to the variables in model 2, we included various information channels (T1: news, television tabloid shows, newspapers, news websites, nonnews websites, and social media) as independent variables in model 3. All variables entered as independent variables for all models were forced into the model based on previous studies or theory as variables associated with information overload.

No missing values were observed because the online survey was designed such that respondents could not proceed to the next question if values were missing. Participants who only responded to T1 were not included in our analysis. All tests were 2-sided, and the significance level was set at 5%. We used IBM SPSS (version 25; IBM Corp) for the analysis.

Results

Participants' characteristics are presented in Table 1. The breakdown of characteristics by sex, age, and prefecture of residence was similar to that of the general population in Japan, as stated in the Methods section. The minimum age was 21 years and the maximum age was 76 years. The proportions of participants with an education level lower than a university degree and with a university degree or higher were similar. Approximately half of the participants were full-time employees. The demographic characteristics of participants who were included in the analysis did not differ statistically from the characteristics of those who were excluded owing to not having completed T2.

Table 1. The characteristics of the study participants (N=784).

Participant characteristics	Values
Age (y), mean (SD)	47.14 (12.8)
Sex, n (%)	
Male	396 (50.5)
Female	388 (49.5)
Education, n (%)	
Junior high school	12 (1.5)
High school	156 (19.9)
Vocational school or junior college	175 (22.3)
University	379 (48.3)
Graduate school	62 (7.9)
Household income, n (%)	
<JP ¥2,000,000 (<US $15,000)	64 (8.2)
JP ¥2,000,000-¥6,000,000 (US $15,000-$43,500)	320 (49)
>JP ¥6,000,000 (>US $43,500)	334 (42.6)
Unclear	66 (8.4)
Previous infection, n (%)	21 (2.7)
Cohabiting with family, n (%)	613 (78.2)
Health literacy[a], mean (SD)	3.822 (0.6)

[a]Health literacy was measured by the Communicative Critical Health Literacy.

Table 2 shows the results of the single regression analysis of sociodemographic characteristics, HL, and attention to media information channels in terms of COVID-19 information overload. Significant simple associations with information overload were found for age, sex, education, HL, television news, television tabloid shows, newspapers, and social media.

XSL•FO
RenderX

Table 2. Univariate associations between information overload and explanatory variables (N=784).

	B	β (SE; 95% CI)	*P* value[a]
Age (y)	−0.006	−0.154 (0.001; −0.008 to −0.003)	<.001
Sex	−0.061	−0.067 (0.033; −0.125 to 0.003)	.06
Education	−0.009	−0.018 (0.017; −0.043 to 0.025)	.61
Household income	−0.052	−0.070 (0.027; −0.104 to 0.000)	.051
Cohabiting with family	−0.057	−0.051 (0.040; −0.134 to 0.021)	.15
Previous infection	−0.286	−0.059 (0.174; −0.627 to 0.054)	.10
Health literacy[b]	−0.082	−0.111 (0.026; −0.133 to −0.030)	.002
Television news	−0.030	−0.186 (0.006; −0.042 to −0.019)	<.001
Television tabloid show	−0.022	−0.137 (0.006; −0.032 to −0.011)	<.001
Newspaper	−0.019	−0.138 (0.005; −0.029 to −0.009)	<.001
News website	−0.017	−0.093 (0.006; −0.030 to −0.004)	.009
Nonnews website	−0.010	−0.048 (0.007; −0.025 to 0.005)	.18
Social media	0.011	0.068 (0.006; 0.000 to 0.021)	.06

[a]The estimates were calculated by simple regression analysis.

[b]Health literacy was measured by the Communicative Critical Health Literacy instrument.

Table 3 shows the results of hierarchical multiple regression analysis. The results of model 1 showed that younger age (β=−0.084, 95% CI −0.142 to −0.013) was significantly more highly correlated with information overload 1 year later. For sex, information overload was significantly higher for men than for women (β=−0.163, 95% CI −0.008 to −0.003). Lower household income was associated with significantly higher information overload (β=−0.083, 95% CI −0.115 to −0.009). The model 2 results showed that the lower the HL, the higher the information overload (β=−0.084, 95% CI −0.114 to −0.011), which was significant. As the results of model 3 show, information overload was significantly higher for those who paid less attention to television news (β=−0.118, 95% CI −0.038 to −0.001) and for those who paid greater attention to social media (β=0.089, 95% CI 0.000-0.027).

XSL•FO
RenderX

Table 3. Multivariate associations between information overload and sociodemographic characteristics, health literacy, and COVID-19 information attention (N=784)[a].

	Model 1			Model 2			Model 3		
	B	β (SE; 95% CI)	*P* value[b]	B	β (SE; 95% CI)	*P* value[b]	B	β (SE; 95% CI)	*P* value[b]
Age (y)	−0.006	−0.158 (0.001; −0.008 to −0.003)	<.001	−0.005	−0.145 (0.001; −0.008 to −0.003)	<.001	−0.003	−0.078 (0.001; −0.006 to 0.000)	.054
Sex	0.076	0.083 (0.033; 0.011 to 0.140)	.022	0.074	0.081 (0.033; 0.009 to 0.138)	.025	0.072	0.079 (0.033; 0.008 to 0.136)	.03
Education	−0.016	−0.033 (0.018; −0.051 to 0.019)	.373	−0.012	−0.026 (0.018; −0.048 to 0.023)	.488	−0.004	−0.008 (0.018; −0.039 to 0.031)	.82
Household income	−0.059	−0.080 (0.029; −0.116 to −0.003)	.040	−0.056	−0.076 (0.029; −0.113 to 0.000)	.051	−0.051	−0.069 (0.029; −0.107 to 0.005)	.08
Cohabiting with family	−0.012	−0.011 (0.042; −0.093 to 0.070)	.778	−0.015	−0.014 (0.042; −0.097 to 0.066)	.714	0.010	0.009 (0.042; −0.072 to 0.092)	.80
Previous infection	−0.221	−0.046 (0.172; −0.559 to 0.117)	.199	−0.209	−0.043 (0.172; −0.546 to 0.128)	.224	−0.210	−0.043 (0.171; 0.546 to 0.127)	.22
Health literacy[c]	—[d]	—	—	−0.062	−0.083 (0.026; −0.113 to −0.01)	.019	−0.061	−0.082 (0.026; −0.112 to 0.009)	.02
Television news	—	—	—	—	—	—	−0.019	−0.117 (0.010; −0.038 to 0.000)	.045
Television tabloid show	—	—	—	—	—	—	0.002	0.011 (0.009; −0.015 to 0.018)	.84
Newspaper	—	—	—	—	—	—	−0.006	−0.042 (0.006; −0.017 to 0.006)	.32
News website	—	—	—	—	—	—	−0.007	−0.036 (0.008; −0.022 to 0.008)	.39
Nonnews website	—	—	—	—	—	—	−0.012	−0.056 (0.009; −0.030 to 0.007)	.22
Social media	—	—	—	—	—	—	0.014	0.089 (0.007; 0.000 to 0.027)	.048

[a]For model 1, R2 was 0.039, ΔR2 was 0.039, $F_{6,777}$=5.292, and *P* value was <.001. For model 2, R2 was 0.046, ΔR2 was 0.007, $F_{1,776}$=5.490, and *P* value was .02. For model 3, R2 was 0.072, ΔR2 was 0.026, $F_{1,770}$=3.532, and *P* value was .002.

[b]The estimates were calculated by hierarchical multiple regression analysis.

[c]Health literacy was measured by the Communicative Critical Health Literacy instrument.

[d]Not applicable.

Discussion

Overview

In this study, we examined predictors of information overload during the COVID-19 state of emergency declaration in Japan. The results suggest that predictors of information overload include individual characteristics, skills for handling health information, and the information channels to which individuals are exposed. These results are important findings for developing public health communication strategies that take into account people who are more likely to perceive information overload during a pandemic.

Individual Characteristics That Predict Information Overload

The results of this study did not support our hypothesis that older people are more likely to experience information overload during a pandemic. There are 2 possible reasons why younger people were more likely to experience information overload during the prolonged COVID-19 pandemic conditions. First, young people had fewer health problems before the pandemic than older people. As a result, younger people may have had less experience searching for and processing health information themselves. Previous studies have reported that people with less experience in information processing are more likely to perceive information overload [12,14]. When appropriate resources are lacking, it becomes difficult to integrate new information with an individual's existing knowledge; therefore, the individual does not know how to react to new information [14]. Second, new media channels, such as social media, have developed rapidly in recent years. Younger people are likely to be exposed to older types of media but also use multiple new media channels [55]. During the COVID-19 pandemic, people were exposed to a greater volume of information from a variety of sources, media, and platforms than they were during previous pandemics [52]. This diversity and complexity of platforms can lead to exposure to irrelevant information and the need to allocate extra

XSL•FO
RenderX

WM [40,48,56]. This may have resulted in a greater cognitive burden on younger people. The results of this study suggest that, in the special information environment of a pandemic, young people may experience greater cognitive load than older people.

Contrary to our hypothesis, men were more likely to perceive information overload than women in this study. Although previous studies have demonstrated sex differences in behaviors related to the COVID-19 pandemic, the relationship between information overload and sex has been unclear [31,36]. During a prolonged pandemic, people are affected psychologically and behaviorally by various external triggers. Similar to information overload, one of these triggers is pandemic fatigue, which is more often perceived by men [57]. Pandemic fatigue has been described by the WHO as a loss of motivation owing to the chronicity of a pandemic situation and reportedly comprises 2 concepts: neglect of preventive behaviors and boredom with overexposure to information [58]. Fatigue from repeated exposure to similar information is reportedly associated with information overload [59]. It is possible that men are more vulnerable than women to fatigue owing to repeated exposure to similar information.

In future pandemics, more diverse information may be disseminated than during the COVID-19 pandemic, and more people may perceive information overload. As the results of this and previous studies have shown, younger people and men may not be able to acquire appropriate and correct knowledge owing to information overload and may have difficulty adopting appropriate preventive behaviors. It is necessary to consider information provision strategies that are appropriate for these populations.

Effect of Health Information–Handling Skills With Information Overload

The results of this study support the hypothesis that people with higher levels of HL are less likely to experience information overload. This finding is consistent with the results of 2 cross-sectional studies conducted in the United States during the COVID-19 pandemic [45,59]. Those results suggest that previous skills in handling health information may reduce the intrinsic cognitive load of complex information in a special information environment, such as a pandemic [40]. The HL scale used in our study measured 2 aspects of HL: communicative literacy, which is the ability to extract information and derive meaning, and critical literacy, which is the ability to critically analyze information. It has been reported that critical thinking is an important skill in mitigating information overload [60]. Critical thinking is closely related to information mastery competencies, such as problem-solving and self-management [61].

As a skill that individuals can develop in advance, HL is an important resource to help combat information overload during a pandemic. HL can be improved through intervention [62]. Therefore, education in developing the skills needed to appropriately process health information should be promoted. Although not addressed in this study, functional literacy is also an important aspect of HL. The variance in functional literacy among Japanese people is small, and it is not suitable for predicting health-related behavior and outcomes. However, additional findings on this aspect will lead to the development of information provision strategies for people with limited functional HL [63].

Information Channels That Predict Information Overload

Regarding the association between the degree of attention to each information channel and information overload, our study results support our hypothesis that greater attention to social media information is a predictor of information overload. In addition, we found that survey respondents who were attentive to television news had a significantly lower perception of information overload. This result can be explained by the characteristics of the information provided by both types of media, as follows. First, the characteristics of the media content differ between these types of media. Television news involves mainly objective facts that have been scrutinized by the news organization, whereas social media information involves mainly subjective perspectives and interpretations by individuals. The 2018 Edelman Trust Barometer Special Report concluded that only 41% of people trust social media [64]. When obtaining information from another individual on social media, there is a cognitive load involved in determining whether that information is worth believing [55]. Concerns about the quality of information may contribute to increasing the cognitive load in information processing [43,65]. A cross-sectional study conducted in the United States during the COVID-19 pandemic reported that individuals receiving information through 2 channels, social media and interpersonal communication, were less likely to perceive information overload than those receiving information through social media alone [65]. Social media information may reduce the cognitive load when provided through interpersonal communication. Second, the originator of information differs among these media sources. Television news is transmitted from a news broadcasting agency by professional news readers, whereas social media information can be transmitted by anyone, including laypersons, and is often put forth by people who are closer and more connected to the individual receiving the information. One cause of the intrinsic cognitive load with social media is FOMO, which can occur when people exchange information among connected individuals [66]. FOMO is defined as a person's experience of pervasive anxiety that others may be having rewarding experiences that they themselves are not experiencing, and is characterized by a desire to continuously connect with what others are doing [67]. This makes it difficult for some individuals to reduce or stop using social media for fear of missing important information and activity on social media [68,69]. From the perspective of cognitive load, personal informational media to which the user has an emotional attachment (eg, social media) may involve a high cognitive load, whereas informational media that provides objective information with no emotional attachment (eg, television news) may produce a lower cognitive load.

On the basis of the results of this research and the findings of previous studies, we propose several approaches to reducing information overload during a pandemic from the perspectives of intrinsic, extrinsic, and germane cognitive load. First, health information about a pandemic is uncertain and changes over

XSL•FO
RenderX

time, which means it has a high intrinsic load. One factor that reduces intrinsic load is previous specialist knowledge and skills. Encouraging people to develop skills, such as HL, information literacy for reading information critically, and slow reading under normal circumstances, will help to reduce the cognitive load caused by the complexity of information during a pandemic [60,70]. Approaches to reducing extrinsic load using technological methods have been proposed. These include methods that improve conciseness, consistency, and understandability of information content, as well as increasing findability, a technique that makes it easier for users to find content based on their needs [60,71]. Finally, for people who have characteristics that make them more susceptible to cognitive overload from complex health information during a pandemic, it may be possible to provide information that is less likely to cause overload by segmentation according to sex and age. These approaches can be adopted as part of an information delivery strategy by public health communicators to mitigate information overload during a pandemic.

Limitations

Several limitations should be considered when interpreting the results of this study. First, because this was an observational study, a causal relationship between individual characteristics or attention to media information and information overload cannot be established from these results. Therefore, we relied on previous studies and existing theories in our discussion. However, given the considerable difficulty in conducting a randomized controlled trial during a public health crisis, the results of this study represent important findings. Second, study participants were registered members of a survey panel recruited by an internet company, and their educational background tended to be higher than that of the general Japanese public. We adjusted for the influence of educational background using multiple regression analysis; however, there is a possibility that there are still unadjusted factors related to self-selection. In addition, considering the R2 scores of each regression model, there may be unknown variables that can better explain information overload. Finally, single-item, unvalidated scales were used in this study to assess the degree of attention to media information. We adopted items that have been commonly used in previous studies to enable comparisons with results from other infectious disease pandemics and findings from other countries. In the future, we anticipate the development of scales that will measure these factors.

Conclusions

During the COVID-19 pandemic, people have been exposed to a great deal of information through multiple media channels, constituting an infodemic. The results of this study have longitudinal implications in terms of the characteristics of individuals and types of informational media that can predict information overload in a specific informational environment. Public health communicators should consider younger people, men, and people with low HL when developing public health communication strategies during a pandemic. High levels of social media attention may increase the perception of information overload. Therefore, public health communicators can incorporate technical approaches to prevent information overload during a pandemic, such as improving the findability of appropriate information and preparing concise, understandable content. Moreover, providing opportunities for educating the public about critical reading of health information during normal periods will help individuals to prepare for future infodemics during which anyone can disseminate information.

Acknowledgments

This work was supported by the Japan Society for the Promotion of Science KAKENHI (grant 20K10397). The authors thank Analisa Avila, MPH, ELS, of Edanz for editing a draft of this paper.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HO and TO contributed to conceptualization and methodology. HO contributed to data curation, formal analysis, and writing the original draft. TK contributed to supervision. HO, TO, and TK contributed to writing, reviewing, and editing.

Conflicts of Interest

None declared.

References

1. Kusama T, Kiuchi S, Takeuchi K, Ikeda T, Nakazawa N, Kinugawa A, et al. Information usage and compliance with preventive behaviors for COVID-19: a longitudinal study with data from the JACSIS 2020/JASTIS 2021. Healthcare (Basel) 2022 Mar 13;10(3):521 [FREE Full text] [doi: 10.3390/healthcare10030521] [Medline: 35326999]
2. Uchibori M, Ghaznavi C, Murakami M, Eguchi A, Kunishima H, Kaneko S, et al. Preventive behaviors and information sources during COVID-19 pandemic: a cross-sectional study in japan. Int J Environ Res Public Health 2022 Nov 04;19(21):14511 [FREE Full text] [doi: 10.3390/ijerph192114511] [Medline: 36361391]
3. No authors listed. The COVID-19 infodemic. Lancet Infect Dis 2020 Aug;20(8):875. [doi: 10.1016/S1473-3099(20)30565-x]

XSL•FO
RenderX

4. Ji Q, Ha L, Sypher U. The role of news media use and demographic characteristics in the possibility of information overload prediction. Int J Commun 2014;8 [FREE Full text]
5. Apuke OD, Omar B. Social media affordances and information abundance: enabling fake news sharing during the COVID-19 health crisis. Health Informatics J 2021 Jul 10;27(3):14604582211021470 [FREE Full text] [doi: 10.1177/14604582211021470] [Medline: 34250868]
6. Balakrishnan V, Ng WZ, Soo MC, Han GJ, Lee CJ. Infodemic and fake news - a comprehensive overview of its global magnitude during the COVID-19 pandemic in 2021: a scoping review. Int J Disaster Risk Reduct 2022 Aug;78:103144. [doi: 10.1016/j.ijdrr.2022.103144] [Medline: 35791376]
7. Lang A. The limited capacity model of mediated message processing. J Commun 2000;50(1):46-70 [FREE Full text] [doi: 10.1111/j.1460-2466.2000.tb02833.x]
8. Miller GA. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev 1956 Mar;63(2):81-97. [doi: 10.1037/h0043158]
9. Malhotra N. Information load and consumer decision making. J Consum Res 1982 Mar;8(4):419-430. [doi: 10.1086/208882]
10. Simon HA. Information processing models of cognition. Annu Rev Psychol 1979 Jan;30(1):363-396. [doi: 10.1146/annurev.ps.30.020179.002051] [Medline: 18331186]
11. Sweller J. Cognitive load during problem solving: effects on learning. Cognit Sci 2010 Feb 11;12(2):257-285. [doi: 10.1207/s15516709cog1202_4]
12. Eppler MJ, Mengis J. The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. Inf Soc 2004 Nov;20(5):325-344. [doi: 10.1080/01972240490507974]
13. Davis TC, Williams MV, Branch WT. Explaining illness to patients with limited literacy. In: Whaley BB, editor. Explaining Illness: Research, Theory, and Strategies. Mahwah, NJ: Lawrence Erlbaum Associates; 1999:171-194.
14. Eppler M. Information quality and information overload: the promises and perils of the information age. In: Cantoni L, Danowski JA, editors. Communication and Technology. Berlin, Germany: De Gruyter Brill; 2015.
15. Jensen JD, Carcioppolo N, King AJ, Bernat JK, Davis L, Yale R, et al. Including limitations in news coverage of cancer research: effects of news hedging on fatalism, medical skepticism, patient trust, and backlash. J Health Commun 2011 May;16(5):486-503 [FREE Full text] [doi: 10.1080/10810730.2010.546491] [Medline: 21347947]
16. Kalyuga S, Ayres P, Chandler P, Sweller J. The expertise reversal effect. Educ Psychol 2010 Jun 08;38(1):23-31. [doi: 10.1207/s15326985ep3801_4]
17. Chae J, Lee CJ, Jensen JD. Correlates of cancer information overload: focusing on individual ability and motivation. Health Commun 2016 Oct 29;31(5):626-634. [doi: 10.1080/10410236.2014.986026] [Medline: 26512760]
18. Liu CF, Kuo KM. Does information overload prevent chronic patients from reading self-management educational materials? Int J Med Inform 2016 May;89:1-8. [doi: 10.1016/j.ijmedinf.2016.01.012] [Medline: 26980353]
19. Chen X, Lee W, Lin F. Infodemic, institutional trust, and COVID-19 vaccine hesitancy: a cross-national survey. Int J Environ Res Public Health 2022 Jun 30;19(13):8033 [FREE Full text] [doi: 10.3390/ijerph19138033] [Medline: 35805691]
20. Lin F, Chen X, Cheng EW. Contextualized impacts of an infodemic on vaccine hesitancy: the moderating role of socioeconomic and cultural factors. Inf Process Manag 2022 Sep;59(5):103013 [FREE Full text] [doi: 10.1016/j.ipm.2022.103013] [Medline: 35874146]
21. Wang J, Huang X, Wang Y, Wang M, Xu J, Li X. COVID-19 information overload, negative emotions and posttraumatic stress disorder: a cross-sectional study. Front Psychiatry 2022 May 13;13:894174 [FREE Full text] [doi: 10.3389/fpsyt.2022.894174] [Medline: 35693965]
22. Guo N, Zhao SZ, Weng X, Wu Y, Luk TT, Wong JY, et al. Associations of COVID-19 online information sources and information overload with psychological distress symptoms: a population-based study. Transl Behav Med 2021 Jul 29;11(7):1330-1338. [doi: 10.1093/tbm/ibab086] [Medline: 34160612]
23. Sweller J. Cognitive load theory. In: Mestre JP, Ross BH, editors. The Psychology of Learning and Motivation: Cognition in Education. Cambridge, MA: Elsevier Academic Press; 2011.
24. Sweller J. Element interactivity and intrinsic, extraneous, and germane cognitive load. Educ Psychol Rev 2010 Apr 23;22(2):123-138. [doi: 10.1007/s10648-010-9128-5]
25. Mathias JL, Lucas LK. Cognitive predictors of unsafe driving in older drivers: a meta-analysis. Int Psychogeriatr 2009 Aug 01;21(4):637-653. [doi: 10.1017/s1041610209009119]
26. Nooyens AC, Wijnhoven HA, Schaap LS, Sialino LD, Kok AA, Visser M, et al. Sex differences in cognitive functioning with aging in the Netherlands. Gerontology 2022 Jan 4;68(9):999-1009. [doi: 10.1159/000520318] [Medline: 34983049]
27. Van Gerven PW, Paas FG, Van Merriënboer JJ, Schmidt HG. Cognitive load theory and aging: effects of worked examples on training efficiency. Learn Instr 2002 Feb;12(1):87-105. [doi: 10.1016/S0959-4752(01)00017-2]
28. Salthouse TA, Babcock RL. Decomposing adult age differences in working memory. Dev Psychol 1991;27(5):763-776. [doi: 10.1037//0012-1649.27.5.763]
29. Wingfield A, Stine EA, Lahar CJ, Aberdeen JS. Does the capacity of working memory change with age? Exp Aging Res 1988 Sep 27;14(2-3):103-107. [doi: 10.1080/03610738808259731] [Medline: 3234452]
30. Hong H, Kim HJ. Antecedents and consequences of information overload in the COVID-19 pandemic. Int J Environ Res Public Health 2020 Dec 12;17(24):9305 [FREE Full text] [doi: 10.3390/ijerph17249305] [Medline: 33322715]

XSL•FO
RenderX

31. Tadiri CP, Gisinger T, Kautzy-Willer A, Kublickiene K, Herrero MT, Raparelli V, et al. The influence of sex and gender domains on COVID-19 cases and mortality. CMAJ 2020 Sep 08;192(36):E1041-E1045 [FREE Full text] [doi: 10.1503/cmaj.200971] [Medline: 32900766]
32. Gender and COVID-19. World Health Organization. 2020 May 14. URL: https://iris.who.int/bitstream/handle/10665/332080/WHO-2019-nCoV-Advocacy_brief-Gender-2020.1-eng.pdf [accessed 2025-08-13]
33. Holton AE, Chyi HI. News and the overloaded consumer: factors influencing information overload among news consumers. Cyberpsychol Behav Soc Netw 2012 Nov;15(11):619-624. [doi: 10.1089/cyber.2011.0610] [Medline: 23020743]
34. York C. Overloaded by the news: effects of news exposure and enjoyment on reporting information overload. Commun Res Reports 2013 Oct;30(4):282-292. [doi: 10.1080/08824096.2013.836628]
35. Sternadori MM, Wise K. Men and women read news differently: the effects of story structure on the cognitive processing of text. J Media Psychol 2010 Jan;22(1):14-25. [doi: 10.1027/1864-1105/a000003]
36. Breyton M, Schultz É, Smith A, Rouquette A, Mancini J. Information overload in the context of COVID-19 pandemic: a repeated cross-sectional study. Patient Educ Couns 2023 May;110:107672 [FREE Full text] [doi: 10.1016/j.pec.2023.107672] [Medline: 36827879]
37. Eysenbach G, Powell J, Kuss O, Sa ER. Empirical studies assessing the quality of health information for consumers on the world wide web: a systematic review. JAMA 2002 May 22;287(20):2691-2700. [doi: 10.1001/jama.287.20.2691] [Medline: 12020305]
38. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care 2008 Oct;53(10):1310-1315. [Medline: 18811992]
39. Friedman DB, Hoffman-Goetz L. An exploratory study of older adults' comprehension of printed cancer information: is readability a key factor? J Health Commun 2007 Jul 20;12(5):423-437. [doi: 10.1080/10810730701438658] [Medline: 17710594]
40. Hollender N, Hofmann C, Deneke M, Schmitz B. Integrating cognitive load theory and concepts of human–computer interaction. Comput Hum Behav 2010 Nov;26(6):1278-1288. [doi: 10.1016/j.chb.2010.05.031]
41. Healthy people 2030. Office of Disease Prevention and Health Promotion, Office of the Assistant Secretary for Health, Office of the Secretary, U.S. Department of Health and Human Services. URL: https://health.gov/healthypeople/priority-areas/health-literacy-healthy-people-2030 [accessed 2025-08-20]
42. Jiang S, Beaudoin CE. Health literacy and the internet: an exploratory study on the 2013 HINTS survey. Comput Hum Behav 2016 May 27;58(1):240-248 [FREE Full text] [doi: 10.1016/j.chb.2016.01.007]
43. Kim K, Lustria M, Burke D, Kwon N. Predictors of cancer information overload: findings from a national survey. Inf Res 2007;12(4) [FREE Full text]
44. Chan YM, Huang H. Weight management information overload challenges in 2007 HINTS: socioeconomic, health status and behaviors correlates. J Consum Health Internet 2013 Apr;17(2):151-167. [doi: 10.1080/15398285.2013.780540]
45. Yamamoto M, Krishnan A, Golden A, Owen G, Schell LM, Mata O, et al. Correlates of COVID-19 information overload and information seeking: evidence from a community survey. Am J Health Promot 2024 May 13;38(4):503-512. [doi: 10.1177/08901171241227302] [Medline: 38217428]
46. Geiger S, Newhagen J. Revealing the black box: information processing and media effects. J Commun 1993;43(4):42-50. [doi: 10.1111/j.1460-2466.1993.tb01303.x]
47. Dhir A, Kaur P, Chen S, Pallesen S. Antecedents and consequences of social media fatigue. Int J Inf Manage 2019 Oct;48:193-202 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2019.05.021]
48. Xiao L, Mou J. Social media fatigue -technological antecedents and the moderating roles of personality traits: the case of WeChat. Comput Hum Behav 2019 Dec;101:297-310. [doi: 10.1016/j.chb.2019.08.001]
49. De la Torre-Díez I, Díaz-Pernas FJ, Antón-Rodríguez M. A content analysis of chronic diseases social groups on Facebook and Twitter. Telemed J E Health 2012;18(6):404-408. [doi: 10.1089/tmj.2011.0227] [Medline: 22650380]
50. Lee AR, Son SM, Kim KK. Information and communication technology overload and social networking service fatigue: a stress perspective. Comput Hum Behav 2016 Feb;55:51-61. [doi: 10.1016/j.chb.2015.08.011]
51. Bala R, Srivastava A, Ningthoujam GD, Potsangbam T, Oinam A, Anal CL. An observational study in Manipur state, India on preventive behavior influenced by social media during the COVID-19 pandemic mediated by cyberchondria and information overload. J Prev Med Public Health 2021 Jan;54(1):22-30 [FREE Full text] [doi: 10.3961/jpmph.20.465] [Medline: 33618496]
52. Sun J, Lee SK. "No more COVID-19 messages via social media, please": the mediating role of COVID-19 message fatigue between information overload, message avoidance, and behavioral intention. Curr Psychol 2023 Jun 02;42(24):1-15 [FREE Full text] [doi: 10.1007/s12144-023-04726-7] [Medline: 37359620]
53. Ishikawa H, Nomura K, Sato M, Yano E. Developing a measure of communicative and critical health literacy: a pilot study of Japanese office workers. Health Promot Int 2008 Sep;23(3):269-274. [doi: 10.1093/heapro/dan017] [Medline: 18515303]
54. Sarkhel S, Bakhla AK, Praharaj SK, Ghosal MK. Information overload regarding COVID-19: adaptation and validation of the cancer information overload scale. Indian J Psychiatry 2020;62(5):481-487 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_974_20] [Medline: 33678827]

XSL•FO
RenderX

55. Wang Y, Dai Y, Li H, Song L. Social media and attitude change: information booming promote or resist persuasion? Front Psychol 2021 Jun 24;12:596071 [FREE Full text] [doi: 10.3389/fpsyg.2021.596071] [Medline: 34248729]
56. Islam AK, Laato S, Talukder S, Sutinen E. Misinformation sharing and social media fatigue during COVID-19: an affordance and cognitive load perspective. Technol Forecast Soc Change 2020 Oct;159:120201 [FREE Full text] [doi: 10.1016/j.techfore.2020.120201] [Medline: 32834137]
57. Jia X, Ahn S, Carcioppolo N. Measuring information overload and message fatigue toward COVID-19 prevention messages in USA and China. Health Promot Int 2023 Jun 01;38(3):daac003 [FREE Full text] [doi: 10.1093/heapro/daac003] [Medline: 35092282]
58. Pandemic fatigue. World Health Organization. 2020. URL: https://iris.who.int/bitstream/handle/10665/335820/WHO-EURO-2020-1160-40906-55390-eng.pdf [accessed 2025-08-20]
59. Mao B, Jia X, Huang Q. How do information overload and message fatigue reduce information processing in the era of COVID-19? An ability–motivation approach. J Inf Sci 2022 Sep 05;50(5):1242-1254. [doi: 10.1177/01655515221118047]
60. Koltay T. The bright side of information: ways of mitigating information overload. J Document 2017 Jul 10;73(4):767-775. [doi: 10.1108/JD-09-2016-0107]
61. Kam JA, Lee CJ. Examining the effects of mass media campaign exposure and interpersonal discussions on youth's drug use: the mediating role of visiting pro-drug websites. Health Commun 2013 Jul 20;28(5):473-485. [doi: 10.1080/10410236.2012.699873] [Medline: 22816432]
62. Sheridan SL, Halpern DJ, Viera AJ, Berkman ND, Donahue KE, Crotty K. Interventions for individuals with low health literacy: a systematic review. J Health Commun 2011 Sep 30;16 Suppl 3(sup3):30-54. [doi: 10.1080/10810730.2011.604391] [Medline: 21951242]
63. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Prom Int 2000;15(3):259-267. [doi: 10.1093/heapro/15.3.259]
64. 2018 Edelman Trust Barometer. Edelman. 2018. URL: https://www.edelman.com/sites/g/files/aatuss191/files/2018-10/2018_Edelman_Trust_Barometer_Global_Report_FEB.pdf [accessed 2025-08-20]
65. Huang Q, Mao B, Jia X, Peng W. COVID-19 information overload mediated the effects of cross-channel information differences on health information elaboration. J Health Commun 2023 Jul 03;28(7):401-411. [doi: 10.1080/10810730.2023.2217097] [Medline: 37232168]
66. Adams SK, Williford DN, Vaccaro A, Kisler TS, Francis A, Newman B. The young and the restless: socializing trumps sleep, fear of missing out, and technological distractions in first-year college students. Int Jt Adolesct Youth 2016 May 11;22(3):337-348. [doi: 10.1080/02673843.2016.1181557]
67. Przybylski AK, Murayama K, DeHaan CR, Gladwell V. Motivational, emotional, and behavioral correlates of fear of missing out. Comput Hum Behav 2013 Jul;29(4):1841-1848. [doi: 10.1016/j.chb.2013.02.014]
68. González-Padilla DA, Tortolero-Blanco L. Social media influence in the COVID-19 pandemic. Int Braz J Urol 2020 Jul;46(suppl.1):120-124 [FREE Full text] [doi: 10.1590/S1677-5538.IBJU.2020.S121] [Medline: 32550706]
69. Shen Y, Zhang S, Xin T. Extrinsic academic motivation and social media fatigue: fear of missing out and problematic social media use as mediators. Curr Psychol 2020 Nov 26;41(10):7125-7131. [doi: 10.1007/s12144-020-01219-9]
70. Bawden D, Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies. J Inf Sci 2008 Nov 21;35(2):180-191. [doi: 10.1177/0165551508095781]
71. Forte A, Andalibi N, Park T, Willever-Farr H. Designing information savvy societies: an introduction to assessability. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI '14; April 26-May 1, 2014; Toronto, ON. [doi: 10.1145/2556288.2557072]

Abbreviations

FOMO: fear of missing out
HL: health literacy
T1: first survey
T2: second survey
WHO: World Health Organization
WM: working memory

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 02.10.24; peer-reviewed by K Masuko, A Xiong; comments to author 10.01.25; revised version received 14.02.25; accepted 30.07.25; published 30.10.25.

Please cite as:
Okada H, Okuhara T, Kiuchi T
Factors Predicting Information Overload During the COVID-19 Pandemic in the Digital Age: Longitudinal Study
Interact J Med Res 2025;14:e67098
URL: https://www.i-jmr.org/2025/1/e67098
doi:10.2196/67098
PMID:

©Hiroko Okada, Tsuyoshi Okuhara, Takahiro Kiuchi. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 30.10.2025. 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 Physicians' Referral Decision-Making for Rehabilitation Outpatient Services in the Health Care Landscape of China: Cross-Sectional Study

Yawei Li[1], OTD; Ruixue Ye[1], PhD; Linlin Shan[1], MM; Kun Wang[1], MM; Kaiwen Xue[1], MM; Zeyu Zhang[1], MM; Yingzi Hao[2], MM; Yucong Zou[1], MM; Xiaoxuan Li[2], MM; Yulong Wang[1], MD

[1]Department of Rehabilitation, Shenzhen Second People's Hospital, The First Affiliated Hospital, Shenzhen University School of Medicine, 3002 Sungang Road, Futian District, Shenzhen, Guangdong, China

[2]School of Rehabilitation Medicine, The Shandong University of Traditional Chinese Medicine, Jinan, China

Corresponding Author:
Yulong Wang, MD
Department of Rehabilitation, Shenzhen Second People's Hospital, The First Affiliated Hospital, Shenzhen University School of Medicine, 3002 Sungang Road, Futian District, Shenzhen, Guangdong, China

Abstract

Background: Stratified health care systems are used globally to optimize medical resource allocation and enhance patient care experiences. Although successfully implemented in countries like the United Kingdom, Australia, and Canada, China's introduction of stratified health care in 2015 has achieved progress in disease management but still faces challenges due to the lack of a comprehensive referral evaluation system and patients' preference for higher-tier medical institutions.

Objective: This study aims to investigate the factors influencing Chinese rehabilitation physicians' referral decisions for outpatient rehabilitation patients. The findings may provide empirical evidence for developing stratified rehabilitation triage tools and constructing a referral evaluation system in China.

Methods: This cross-sectional study, conducted from September 2023 to January 2024, examined the patient factors (diagnosis, functional impairments, disease status, condition stability, duration of illness, and functional status measured via the Longshi Scale) impacting physicians' referral decisions for outpatient rehabilitation services in China. Data were collected through convenient stratified sampling from physicians and outpatient rehabilitation patients across 12 medical institutions in 5 cities in China.

Results: A total of 131 rehabilitation physicians conducted diversion assessments for 1984 outpatient rehabilitation patients in this study. In total, 45.5% (902/1984) of outpatient rehabilitation patients were considered by physicians to be referred to rehabilitation outpatient clinics, 19% (376/1984) to primary health care institutions, 20.4% (405/1984) to secondary institutions, and 15.2% (301/1984) to tertiary institutions. Single-factor analysis indicated that age, disease, functional impairment, disease control, disease stability, and Longshi Scale results were significantly associated with physicians' decisions regarding the referral institutions for outpatient rehabilitation patients. Logistic regression analysis showed that neurological disorders (odds ratio [OR] 1.88, 95% CI 1.02 - 3.43; *P*=.04), cardiopulmonary diseases (OR 2.91, 95% CI 1.07 - 7.93; *P*=.04), geriatric conditions (OR 0.40, 95% CI 0.23 - 0.68; *P*<.001), disease control (OR 0.23, 95% CI 0.13 - 0.34; *P*<.001), and Longshi Scale results for the bedridden (OR 0.10, 95% CI 0.14 - 0.34; *P*<.001), and domestic groups (OR 0.24, 95% CI 0.14 - 0.34; *P*<.001) as independent factors for referrals to tertiary versus primary institutions. Orthopedic diseases (OR 3.27, 95% CI 1.89 - 5.67; *P*<.001), geriatric conditions (OR 0.58, 95% CI 0.33 - 1.87; *P*=.009), cognitive impairments (OR 1.98, 95% CI 1.17 - 3.36; *P*=.01), multiple impairments (OR 0.35, 95% CI 0.18 - 0.70; *P*=.002), and disease control (OR 0.26, 95% CI 0.15 - 0.37; *P*<.001) were key factors for tertiary versus secondary referrals.

Conclusions: In advancing China's rehabilitation triage in the future, gaining a deep understanding of the key factors influencing physicians' decisions and quickly establishing a referral evaluation system will facilitate the accurate diversion of outpatient rehabilitation patients, enabling them to receive convenient, high-quality, and low-cost medical services. In addition, it will assist the government in reasonably and effectively allocating medical resources, thus achieving the optimization and coordination of the health care system.

(Interact J Med Res 2025;14:e64464) doi:10.2196/64464

KEYWORDS

rehabilitation hierarchical medical system; medical institutions; influencing factors; referrals; outpatient; rehabilitation; decision-making; mobile phone

XSL•FO
RenderX

Introduction

Stratified health care systems have garnered significant attention globally as a pivotal strategy to optimize medical resource allocation and enhance patient care experiences. Countries like the United Kingdom, Australia, and Canada have implemented stratified triage models within their health care frameworks, showcasing successful applications in improving medical efficiency and patient satisfaction [1-3]. These systems prioritize timely treatment for emergency cases while streamlining nonemergency care, thereby alleviating the burden on health care resources and ensuring equitable access to quality care across diverse geographic regions [4,5]. However, despite the widespread adoption of stratified health care models internationally, challenges persist, including the need for standardized patient referral evaluation systems and addressing patient preferences for seeking care at higher-tier facilities [6,7]. Stratified health care, also known as hierarchical medical care, is a system designed to optimize the allocation of medical resources by categorizing health care delivery into primary, secondary, and tertiary levels based on disease severity, treatment needs, and institutional capacity. The core principles of stratified health care include graded diagnosis and treatment, bidirectional referrals, and collaborative care among different health care tiers. In the context of rehabilitation services, stratified health care plays a critical role in ensuring that patients with varying levels of functional impairment and disease complexity receive appropriate and timely interventions. By directing routine or stable cases to community-based institutions and reserving higher-level hospitals for complex, resource-intensive rehabilitation needs, this model enhances care efficiency, reduces medical costs, and improves patient outcomes. The implementation of stratified health care systems is especially important in aging populations with increasing demand for long-term rehabilitation and chronic disease management.

In China, notable progress has been made in advancing stratified health care initiatives since the State Council introduced the concept of stratified health care in 2015 [8]. This approach, characterized by "primary care as the entry point, bidirectional referral, classification of urgency and seriousness, and coordination between upper and lower levels," has demonstrated considerable potential in improving the management of common diseases like stroke [9]. However, the absence of a comprehensive patient referral evaluation system poses challenges to informed decision-making among health care providers within the stratified health care framework. Furthermore, patient preferences for seeking care at higher-tier facilities remain a barrier to the effective implementation of stratified health care, hindering the equitable distribution of medical resources [10].

Within the context of stratified health care, the decision-making process for outpatient rehabilitation patient referrals by physicians plays a critical role in optimizing resource use and enhancing patient care experiences [11]. Yet, there is a paucity of research exploring the factors influencing physician referral decisions for rehabilitation patients across different health care institutions. Therefore, this study aims to investigate the factors influencing Chinese rehabilitation physicians' referral decisions for outpatient rehabilitation patients, with the findings expected to provide preliminary insights for optimizing referral processes in China's rehabilitation system or similar health care systems worldwide.

The decision to focus specifically on outpatient rehabilitation patients was driven by the unique referral challenges and service demands associated with this population. Unlike acute care patients, rehabilitation patients often present with chronic or progressive conditions requiring continuous, multidisciplinary management across institutional levels. Their care pathways are more dependent on timely and appropriate referral decisions to ensure functional recovery and long-term quality of life. Furthermore, rehabilitation services are increasingly being decentralized to primary and community-level institutions in line with health system reforms, making referral decisions a critical juncture for aligning patient needs with institutional capabilities. Despite this, existing literature rarely explores how physicians determine the most suitable rehabilitation setting within stratified health care frameworks, leaving a significant gap in understanding referral behavior specific to this domain.

Methods

Ethical Considerations

This study received ethical approval from the Clinical Research Ethics Committee of the Second People's Hospital of Shenzhen (2023-226-02PJ). Electronic informed consent was obtained from all participants before their inclusion in the study. All collected data were anonymized to protect participant privacy. Personal identifiers were removed, and data were stored securely in password-protected systems accessible only to the research team. Participants received no financial compensation for their time and participation.

Study Design

This study used a cross-sectional design to investigate the factors influencing physicians' referral decision-making for outpatient rehabilitation services within the Chinese health care context. Data were collected from September 2023 to January 2024 across 12 medical institutions in 5 cities, which were Shenzhen (Guangdong Province), Linfen (Shandong Province), Chengdu (Sichuan Province), Haikou (Hainan Province), and Changzhou (Jiangsu Province). A convenient stratified sampling method was used to recruit physicians and outpatient rehabilitation patients.

The term "stratified sampling" in this study refers to the selection of medical institutions across multiple administrative levels and geographic regions to ensure representativeness and diversity in the data. Specifically, stratification was conducted based on two key dimensions: (1) institutional tier (primary, secondary, and tertiary health care institutions) and (2) geographic location, involving 5 representative cities across different provinces (Shenzhen, Linfen, Chengdu, Haikou, and Changzhou). Within each stratum, a convenience sampling approach was applied to recruit rehabilitation physicians and patients who met the inclusion criteria. This stratified strategy aimed to capture variations in physician referral decision-making

XSL•FO
RenderX

across different health care settings and regional contexts, thereby enhancing the generalizability and validity of the study findings.

The 12 participating medical institutions in this study were selected to reflect a diverse range of health care settings in China. These included 4 tertiary general hospitals, 3 secondary-level regional hospitals, and 5 primary-level community health centers or township hospitals. All institutions were publicly funded, aligning with the dominant structure of China's public health care system. Tertiary hospitals included urban academic medical centers with comprehensive rehabilitation departments and multidisciplinary teams. Secondary hospitals were typically municipal or district-level facilities offering intermediate rehabilitation services, while primary institutions focused on basic rehabilitative care and chronic condition management. This institutional diversity allowed for the assessment of referral behaviors across different tiers of medical care and ensured that the findings could be generalized to a broad spectrum of health care providers within the Chinese public health system.

In the context of China's hierarchical health care system, medical institutions are categorized into primary, secondary, and tertiary levels based on their scale, service scope, and technical capabilities. Primary health care institutions, such as community health service centers and township health clinics, primarily provide basic medical care, health education, and rehabilitation services for common and chronic diseases, focusing on prevention and routine functional maintenance. Secondary health care institutions, typically regional hospitals, offer more specialized rehabilitation services, including early-stage post-discharge rehabilitation, physical therapy, and occupational therapy, and serve as referral hubs between primary and tertiary levels. Tertiary health care institutions are large general or specialized hospitals equipped with advanced medical technology and multidisciplinary rehabilitation teams. They provide comprehensive and intensive rehabilitation services for patients with complex, severe, or multiple comorbid conditions, including postsurgical recovery, neurological rehabilitation, and interdisciplinary care. These tiered definitions guided the physician referral decisions analyzed in this study.

Study Population

Physicians included in the study were those actively engaged in referring patients to rehabilitation outpatient services within the selected medical institutions. Patients eligible for participation were those seeking rehabilitation outpatient services during the study period. The study participants consist of licensed physicians from relevant institutions who possess clinical experience and voluntarily participate in the research, as well as outpatient rehabilitation patients. The inclusion criteria for patients are as follows: (1) in need of rehabilitation therapy during hospitalization, (2) able to complete questionnaires cooperatively, (3) outpatient status, and (4) voluntary participation in the study.

In the Chinese medical system, physicians' professional titles reflect their level of clinical expertise, academic qualifications, and years of practice and are typically categorized into 3 levels. Entry-level titles (eg, Resident Physician or Physician-in-Training) usually indicate junior practitioners with fewer years of clinical experience and who may still be under supervision. Intermediate-level titles (eg, Attending Physician or Physician) are held by those with several years of independent clinical practice, typically responsible for routine diagnosis, treatment, and initial referral decisions. Senior-level titles (eg, Chief Physician or Associate Chief Physician) denote highly experienced practitioners who often serve in leadership roles, participate in policy development, and handle complex cases requiring multidisciplinary expertise. These distinctions in professional title were considered during the data collection process and may potentially influence physician decision-making behavior.

The required sample size was determined through an a priori power analysis based on our primary hypothesis that physician referral patterns differ significantly across institutional tiers. The proportions from 3 groups were compared in our study, thus "Tests for Multiple Proportions" were used to calculate the total sample size. A sample of 1404 subjects, divided among 3 groups, achieves a power of 0.90. This power assumes a chi-squared test from an analysis with 2 degrees of freedom is used at a significance level of 0.05. The group proportions under the alternative hypothesis are 0.5, 0.6, and 0.6. The effect size is 0.07 [12-15]. The final patient sample in our study (n=1984) exceeded requirements.

Data Collection

This study used a customized survey embedded within the WeChat Longhai Rehabilitation Mini Program to gather data. The survey content for this study includes 2 parts: the physician questionnaire and the patient questionnaire. The physician questionnaire covers the physician's gender, age, education level, specialty, professional title, and years of work experience. The patient questionnaire includes the patient's gender, age, and medical condition, including the reason for the current visit, the Longshi Scale assessment, the course of the disease, whether there is functional impairment, whether the condition is stable, and the physician's diversion recommendations. The Longshi Scale, designed by the Rehabilitation Medicine Department team at the Shenzhen Second People's Hospital, is a graphical scale for assessing the daily living abilities of patients with functional impairments. It has been shown to have good reliability and validity [16]. Participating physicians voluntarily completed a preproject questionnaire to provide their personal information before the commencement of the study. In addition, participating physicians received training on the Longshi Scale to ensure proficiency in its assessment. After the initiation of the project, physicians collected patient questionnaires from eligible individuals to obtain patient characteristic information and to assess the Longshi Scale. Furthermore, physicians recorded their recommendations regarding appropriate medical institutions for patient diversion.

Data were collected using the WeChat Longhai Rehabilitation Mini Program, a secure, mobile-based health data platform embedded within China's widely used WeChat app (Tencent). The program was jointly developed by the Rehabilitation Medicine Department of Shenzhen Second People's Hospital and a certified medical IT provider. It is designed to facilitate

XSL•FO
RenderX

clinical data entry, patient self-assessment, and rehabilitation tracking in both hospital and community settings. For this study, a customized electronic questionnaire was integrated into the platform to allow participating physicians to complete assessments and record patient data in real time using their smartphones or tablets. The platform includes built-in data encryption and access control features to ensure data privacy and security, and all data were stored on a password-protected hospital server compliant with Chinese medical data protection regulations. Before the study, physicians received standardized training on using the platform and conducting assessments to ensure uniformity in data collection. The training included video demonstrations, practice cases, and a short quiz to assess comprehension. Physicians had to pass this quiz before proceeding to data collection.

The data collection link for the physician questionnaire, embedded within the WeChat Longhai Rehabilitation Mini Program, was distributed through institutional communication channels. Specifically, the research coordination team at each participating hospital sent the link to eligible physicians via official WeChat workgroups, which are commonly used in Chinese clinical settings for internal communication. The invitation message included a brief description of the study, an ethics approval notice, and instructions for voluntary participation. The link was shared up to 3 times over a 2-week recruitment period to ensure adequate response rates without exerting undue pressure.

To prevent multiple submissions from the same participant, the survey platform incorporated a real-name login system requiring physicians to enter their unique employee ID number and name, which were verified against institutional staff lists. For patients, duplicate submissions were prevented by requiring patient IDs during registration, which were cross-verified with hospital records. The system automatically flagged and rejected multiple submissions under the same patient ID, ensuring only 1 response per participant was recorded. In addition, the system restricted each ID to 1 completed submission, and duplicate entries were automatically blocked. Data consistency and completeness were reviewed weekly by the research team to detect and address any anomalies in the submission records.

Before formal data collection, the complete questionnaire—including both the physician and patient components—was pilot tested among a small sample of participants to assess clarity, relevance, and usability. Specifically, 5 rehabilitation physicians and 15 outpatient rehabilitation patients from a tertiary hospital not included in the main study were invited to complete the survey via the WeChat Longhai Rehabilitation Mini Program. Feedback was collected on the comprehensibility of the items, the logic and flow of the electronic form, and the usability of the platform interface. Minor modifications were made based on this feedback, including clarifying the wording of several items and optimizing the layout for mobile devices. The pilot phase confirmed the feasibility of using the digital survey format and supported the content validity of the questionnaire before its broader deployment across study sites.

Variable Measurement

The variables of interest encompassed a broad spectrum of factors potentially influencing physicians' referral decisions. These included demographic characteristics of physicians, patient demographics, clinical factors, and institutional characteristics. Variables were carefully defined and operationalized to ensure consistency and reliability across data collection instruments. By comparing personalized patient health care–seeking factors with physicians' experiences, key influencing factors in the decision-making process were analyzed.

The primary outcome was self-reported referral decisions by physicians (including referral to tertiary, secondary, or primary institutions), which was collected by the Longhai Rehabilitation Mini Program in the WeChat app. Exposures included: (1) Physician characteristics, including demographics (gender, age categorized as <40, 40 - 50, and >50 y), work setting (primary, secondary, and tertiary health care institution), education level (college, undergraduate, master, and doctor), professional title (entry-level, intermediate-level, and senior-level), clinical experience (1 - 5, 6 - 10, 11 - 20, and >20 y), and specialty (Western medicine, traditional Chinese medicine, integrative medicine), were collected via self-reported electronic questionnaires; (2) Patient factors were assessed by trained physicians using standardized electronic questionnaires. The assessment included patient demographics (gender, age categorized as <40, 40 - 60, and >60 y), clinical diagnosis (orthopedic, neurological, cardiopulmonary, geriatric, and cancer disease), functional impairment (cognitive, speech, swallowing, cardiopulmonary, motor, and multiple functional impairments), disease status (controlled or uncontrolled), vital signs (stable or unstable), illness duration (less than 12 mo or more than 12 mo), Longshi Scale (bedridden, domestic, or community group). It should be noted that these findings reflect physicians' referral intentions rather than actual patient follow-through, as China's tiered health care system implementation remains incomplete.

The Longshi Scale is a graphical assessment tool used to evaluate the daily living abilities of patients with functional impairments and is widely applied in rehabilitation settings in China. It categorizes patients into three functional levels based on their capacity for self-care and participation in daily activities: (1) the community group, in which patients are able to perform basic daily tasks independently and participate in community or social activities; (2) the domestic group, in which patients are able to care for themselves partially and perform some domestic tasks but are mostly limited to the home environment; and (3) the bedridden group, in which patients are fully dependent on caregivers for mobility and daily care and are confined to bed for most activities.

In this study, physicians used the Longshi Scale to classify patients into one of these three groups, which served as a proxy for functional impairment severity in the regression analysis.

Statistical Analysis

All statistical analyses were conducted using SPSS (version 26.0; IBM Corp). Descriptive statistics were used to summarize physician and patient characteristics. Categorical variables were

XSL•FO
RenderX

expressed as frequencies and percentages, and continuous variables were presented as mean (SD) after normality testing using the Shapiro–Wilk test. Group comparisons for categorical variables were assessed using χ^2 tests or Fisher exact tests where appropriate.

For bivariate analysis, Pearsons χ^2 test was used to assess associations between patient characteristics and physician referral decisions across health care levels. Variables that showed significant associations (P<.05) in univariate analyses were entered into a multinomial logistic regression model to identify independent predictors of referral destination (tertiary vs primary or secondary institutions).

Before regression modeling, multicollinearity was assessed using the variance inflation factor (VIF), and no significant collinearity was detected (all VIFs <2). The assumption of independence of errors was tested using the Durbin-Watson statistic, and model goodness-of-fit was evaluated using the likelihood ratio test and pseudo R^2 values (Nagelkerke).

Confounding control was performed by simultaneously including clinically and statistically relevant variables in the multivariate model based on literature review and bivariate significance. Subgroup analyses were planned based on disease type (eg, orthopedic, neurological) and functional impairment status. However, due to limitations in sample size for certain subgroups, interaction effects between key predictors (eg, disease type×Longshi Scale group) were not formally tested in this version of the analysis and will be considered in future studies.

Missing data were minimal (<5%) and assumed to be missing at random. Listwise deletion was applied in multivariate analyses, and sensitivity checks confirmed that the exclusion of cases with missing values did not significantly alter the main findings.

A 2-tailed P value of less than .05 was considered statistically significant.

Results

Characteristics of Participants

Initially, a total of 148 rehabilitation physicians and 2117 outpatient rehabilitation patients were screened for inclusion. After applying eligibility criteria, 17 physicians were excluded due to insufficient involvement in referral decision-making, incomplete demographic data, or lack of informed consent. Similarly, 133 patients were excluded based on the following reasons: inpatient status during recruitment (n=54), incomplete questionnaires (n=41), inability to provide informed consent due to severe cognitive or communication impairments without a caregiver (n=26), or concurrent participation in conflicting research protocols (n=12). The final analysis thus included 131 physicians and 1984 patients who met all inclusion criteria and completed the required assessments.

Table 1 summarizes the demographic characteristics of physicians. This study included 131 physicians, comprising 57 males (43.5%) and 74 females (56.5%). Regarding age distribution, there were 86 physicians aged 40 years and younger (65.6%), 39 aged between 40 and 50 years (29.8%), and 6 aged 50 years and older (4.6%). Among them, 51 worked in tertiary medical institutions (38.9%), 36 in secondary institutions (27.5%), and 44 in primary institutions (33.6%). In terms of educational background, 1 had a doctoral degree (0.8%), 32 had master's degrees (24.4%), 94 had bachelor's degrees (71.8%), and 4 had associate degrees (3.1%). There were 51 physicians with senior professional titles (38.6%), 36 with intermediate titles (27.5%), and 44 with junior titles (33.9%). Regarding work experience, 17 physicians had 1 - 5 years of experience (13%), 35 had 6 - 10 years of experience (26.7%), 52 had 10 - 20 years of experience (39.7%), and 27 had over 20 years of experience (20.9%). Please see Table 1 for details.

This study enrolled a total of 1984 outpatient rehabilitation patients. Among them, 376 patients (19%) opted for primary health care institutions, 405 patients (20.4%) for secondary health care institutions, 301 patients (15.2%) for tertiary health care institutions, and 902 patients (45.5%) attended rehabilitation outpatient clinics. Of all patients, 1122 were male (56.6%) and 862 were female (43.4%). Regarding age distribution, 39.4% (781/1984) were ages 40 years and younger, 29.3% (582/1984) were between ages 40 and 60 years, and 31.3% (621/1984) were ages 60 years and older. Diagnosis-wise, 41.6% (825/1984) had orthopedic conditions, 27.2% (549/1984) neurological disorders, 4.2% (84/1984) cardiopulmonary diseases, 12% (238/1984) geriatric conditions, and 1.2% (24/1984) had cancer. In terms of functional impairments, 31.8% (631/1984) had cognitive impairments, 30.4% (603/1984) had speech impairments, 9.2% (183/1984) had swallowing difficulties, 6.2% (123/1984) had cardiopulmonary issues, 65.7% (1304/1984) had motor impairments, 19.5% (/1984) had other impairments, and 37.9% (752/1984) exhibited multiple functional impairments. Disease control was inadequate in 14.5% (288/1984) of patients, and 1% (19/1984) exhibited unstable vital signs. Patients with a disease duration of less than 12 months accounted for 82.5% (1637/1984). According to the Longshi Scale assessment, 23.7% (470/1984) were bedridden, 16.6% (329/1984) required family care, and 59.7% (1185/1984) were integrated into social life. Table 2 summarizes the demographic characteristics of patients.

XSL•FO
RenderX

Table . Physician demographic information (N=131).

Characteristics of physicians	Frequency, n (%)
Sex	
Male	57 (43.5)
Female	74 (56.5)
Age (years)	
<40	86 (65.6)
40 - 50	39 (29.8)
>50	6 (4.6)
Work setting	
Primary health care institution	44 (33.6)
Secondary health care institution	36 (27.5)
Tertiary health care institution	51 (38.9)
Experience (years)	
1 - 5	17 (13)
6 - 10	35 (26.7)
11 - 20	52 (39.7)
More than 20	27 (20.6)
Education level	
College	4 (3.1)
Undergraduate	94 (71.8)
Master	32 (24.4)
Doctor	1 (0.8)
Professional title	
Entry-level	44 (33.6)
Intermediate-level	36 (27.5)
Senior-level	51 (38.9)
Specialty	
Western medicine	45 (34.4)
Traditional Chinese medicine	64 (48.9)
Integrative medicine	16 (12.2)

XSL•FO
RenderX

Table . Single-factor analysis of physicians' decisions regarding patient referrals.

Factors	Total (N=1984)	Level of medical institutions			Chi-square (*df*)	*P* value
		Primary (n=376)	Secondary (n=405)	Tertiary (n=301)		
Sex					28.41 (3)	<.001
Male	1122 (56.6)	182 (9.2)	238 (12)	146 (7.4)		
Female	862 (43.4)	194 (9.8)	167 (8.4)	155 (7.8)		
Age (years)					548.89 (6)	<.001
<40	781 (39.4)	85 (4.3)	63 (3.2)	43 (2.3)		
40 - 60	582 (29.3)	155 (7.8)	124 (6.2)	96 (4.8)		
>60	621 (31.3)	136 (6.7)	218 (11.0)	62 (3.1)		
Diagnosis						
Orthopedic					137.76 (3)	<.001
No	1159 (58.4)	122 (6.1)	283 (14.3)	177 (8.9)		
Yes	825 (41.6)	254 (12.9)	122 (6.1)	124 (6.3)		
Neurological					495.40 (3)	<.001
No	1444 (72.8)	295 (14.9)	172 (8.7)	137 (6.9)		
Yes	549 (27.2)	81 (4.1)	233 (11.7)	164 (8.3)		
Cardiopulmonary					68.14 (3)	<.001
No	1900 (95.8)	368 (18.6)	370 (18.6)	271 (13.7)		
Yes	84 (4.2)	8 (0.4)	35 (1.8)	30 (1.5)		
Geriatric					101.80 (3)	<.001
No	1746 (88)	311 (15.7)	317 (16)	254 (12.8)		
Yes	238 (12)	65 (3.3)	88 (4.4)	47 (2.4)		
Cancer					10.31 (3)	.02
No	1960 (98.8)	370 (18.7)	395 (19.9)	297 (15)		
Yes	24 (1.2)	6 (0.3)	10 (0.5)	4 (0.2)		
Functional impairments						
Cognitive impairment					128.38 (3)	<.001
No	1353 (68.2)	346 (17.5)	274 (13.8)	189 (9.5)		
Yes	631 (31.8)	30 (1.5)	131 (6.6)	112 (5.7)		
Speech impairment					114.23 (3)	<.001
No	1381 (69.6)	347 (17.5)	262 (13.2)	185 (9.3)		
Yes	603 (30.4)	29 (1.5)	143 (7.2)	116 (5.9)		
Swallowing impairment					242.09 (3)	<.001
No	1801 (90.8)	366 (18.5)	332 (16.7)	217 (10.9)		
Yes	183 (9.2)	10 (0.5)	73 (3.7)	84 (4.2)		
Cardiopulmonaryimpairment					148.34 (3)	<.001
No	1861 (93.8)	363 (18.3)	350 (17.6)	251 (12.7)		
Yes	123 (6.2)	13 (0.7)	55 (2.8)	50 (2.5)		

XSL•FO
RenderX

Factors	Total (N=1984)	Level of medical institutions			Chi-square (*df*)	*P* value
		Primary (n=376)	Secondary (n=405)	Tertiary (n=301)		
Motor Impairment					552.8 (3)	<.001
No	680 (34.3)	65 (3.3)	47 (2.4)	14 (0.7)		
Yes	1304 (65.7)	311 (15.7)	358 (18)	287 (14.5)		
Multiple functional impairments					147.03 (3)	<.001
No	1232 (62.1)	320 (16.1)	185 (9.3)	154 (7.8)		
Yes	752 (37.9)	56 (2.8)	220 (11.1)	147 (7.4)		
Disease status uncontrolled					75.36 (3)	<.001
No	288 (14.5)	57 (2.9)	45 (2.3)	91 (4.6)		
Yes	1696 (85.5)	319 (16.1)	360 (18.1)	210 (10.6)		
Unstable vital signs					44.01 (3)	<.001
No	19 (1)	0 (0)	7 (0.4)	12 (0.6)		
Yes	1965 (99)	376 (19)	398 (20)	289 (14.6)		
Duration of illness less than 12 months					64.91 (6)	<.001
No	346 (17.4)	104 (5.2)	75 (3.8)	69 (3.5)		
Yes	1637 (82.5)	272 (13.7)	329 (16.6)	232 (11.7)		
Longshi Scale					641.93 (6)	<.001
Community group	1185 (59.7)	283 (14.3)	107 (5.4)	63 (3.2)		
Domestic group	329 (16.6)	52 (2.6)	116 (5.8)	63 (3.2)		
Bedridden group	470 (23.7)	41 (2.1)	182 (9.2)	175 (8.8)		

Single-Factor Analysis of Physicians' Decisions Regarding Patient Referrals

Patient age, sex, and diagnosis type significantly influence physician decisions, including orthopedic, neurological, cardiopulmonary, and geriatric conditions. Functional impairments such as cognitive, speech, swallowing, cardiopulmonary, and motor impairments also exhibit significant differences in physician decision-making (*P*<.001). In addition, factors such as multiple functional impairments, disease control status, stable vital signs, and assessments using the Longshi Activities of Daily Living (ADL) scale significantly impact physician triage decisions (*P*<.001). See Table 2 for detailed findings.

Multinomial Logistic Regression Analysis of Factors Influencing Physicians' Decisions on Patient Referral Destinations

The results in Table 3 represent the multivariate regression analysis comparing the triage decisions of tertiary health care institutions with primary and secondary health care institutions, respectively.

Comparing the triage decision results between primary and tertiary healthcare institutions, neurological disorders (odds ratio [OR] 1.88, 95% CI 1.02 - 3.43; *P*=.04), cardiopulmonary diseases (OR 2.91, 95% CI 1.07 - 7.93; *P*=.04), geriatric conditions (OR 0.40, 95% CI 0.23 - 0.68; *P*<.001), disease uncontrol (OR 0.23, 95% CI 0.13 - 0.34; *P*<.001), and Longshi Scale assessment results for bedridden group (OR 0.10, 95% CI 0.14 - 0.34; *P*<.001) and domestic group (OR 0.24, 95% CI 0.06 - 0.18; *P*<.001) were identified as independent influencing factors for physician decision-making on the destination of rehabilitation outpatients (*P*<.05).

Comparing the triage decision results between secondary and tertiary health care institutions, factors such as orthopedic diseases (OR 3.27, 95% CI 1.89 - 5.67; *P*<.001), geriatric conditions (OR 0.58, 95% CI 0.33 - 1.87; *P*=.009), cognitive impairments (OR 1.98, 95% CI 1.17 - 3.36; *P*=.01), multiple functional impairments (OR 0.35, 95% CI 0.18 - 0.70; *P*=.002), and disease uncontrolled (OR 0.24, 95% CI 0.15 - 0.37; *P*<.001) were identified as independent influencing factors for physician decision-making on the destination of rehabilitation outpatients (*P*<.05).

XSL•FO
RenderX

Table . Multinomial logistic regression analysis of factors influencing physicians' decisions on patient referral destinations.

Variables	Model 1[a,b] (primary vs tertiary[d])		Model 2[c,b] (secondary vs tertiary[d])	
	OR[e] (95% CI)	*P* value	OR (95% CI)	*P* value
Diagnosis				
No[d]	Reference	Reference	Reference	Reference
Orthopedic	1.75 (0.97 - 3.16)	.06	3.27 (1.89 - 5.67)	<.001
Neurological	1.88 (1.02 - 3.43)	.04	1.59 (0.93 - 2.71)	.09
Cardiopulmonary	2.91 (1.07 - 7.93)	.04	1.38 (0.70 - 2.71)	.35
Geriatric	0.40 (0.23 - 0.68)	<.001	0.58 (0.33 - 1.87)	.009
Functional impairment				
No[d]	Reference	Reference	Reference	Reference
Cognitive impairments	1.73 (0.77 - 3.83)	.19	1.98 (1.17 - 3.36)	.01
Multiple functional impairments				
No[d]	Reference	Reference	Reference	Reference
Yes	1.48 (0.58 - 3.75)	.41	0.35 (0.18 - 0.70)	.002
Disease status uncontrolled				
No[d]	Reference	Reference	Reference	Reference
Yes	0.23 (0.13 - 0.34)	<.001	0.24 (0.15 - 0.37)	<.001
Longshi Scale				
Community group[d]	Reference	Reference	Reference	Reference
Domestic group	0.24 (0.06 - 0.18)	<.001	0.80(0.48 - 1.31)	.37
Bedridden group	0.10 (0.14 - 0.34)	<.001	0.47(0.28 - 0.77)	.003

[a]Model 1 represents the comparison results between primary care institutions and tertiary healthcare institutions.

[b]Both the above model 1 and model 2 control the influence of sex, age, education, professional title, working years, hospital type, and specialty of doctors.

[c]Model 2 represents the comparison results between secondary care institutions and tertiary healthcare institutions.

[d]Control group.

[e]OR: odds ratio.

Discussion

Principal Findings

This study was designed to identify key clinical and functional factors that influence physician referral decisions for outpatient rehabilitation patients in the context of China's hierarchical health care system. By systematically analyzing physician and patient data across multiple institutional levels and regions, the study aimed to inform the development of stratified referral evaluation tools. The findings directly address this aim by highlighting that disease type (eg, neurological and orthopedic), functional impairment (eg, cognitive and multifunctional), disease control status, and Longshi Scale classification are independently associated with referral destinations. These results confirm the hypothesis that physician referral behavior is shaped not only by institutional factors but also by patient complexity and functional needs. Furthermore, the study adds to the current literature by quantifying these influences through multinomial regression and suggesting potential criteria for standardizing rehabilitation referral decisions.

The results of the multinomial logistic regression analysis offer valuable insights into the determinants influencing physicians' decisions regarding the referral destinations of rehabilitation outpatients. These findings underscore the importance of specific medical conditions and functional impairments in the decision-making process for patient triage between different levels of healthcare institutions.

This study reveals significant differences in referral decisions for orthopedic, neurological, cardiopulmonary, and geriatric diseases. In particular, when comparing tertiary with primary and secondary health care institutions, patients with geriatric diseases and poor disease control are more likely to be referred to higher-level tertiary institutions. In contrast, those with neurological or cardiopulmonary diseases are more likely to be referred to secondary institutions. This may reflect the fact that geriatric patients may face more health problems and complex rehabilitation needs, prompting physicians to refer them to

XSL•FO
RenderX

higher-level medical institutions for more specialized and comprehensive treatment [17]. Similarly, patients with poor disease control require more extensive resources and specialized medical teams [18].

The complexity and specificity of different diseases in rehabilitation also influence referral strategies. Professional referral is a key factor because orthopedic diseases typically require specialized orthopedic physicians and equipment for treatment and rehabilitation. Therefore, physicians may be more inclined to refer these patients to secondary institutions that offer the necessary specialized orthopedic services and resources [19]. In addition, cardiopulmonary diseases are often chronic and require long-term management and monitoring. Patients with cardiopulmonary diseases and functional impairments are more likely to be referred to secondary institutions because these facilities are better suited for managing chronic diseases, integrating community medical resources, providing specialized services for the elderly, and meeting comprehensive rehabilitation needs [20].

The study also found that patients with cognitive impairments are more likely to be referred to secondary institutions. Patients with cognitive impairments, motor impairments, and multiple functional disabilities are more likely to be referred to secondary institutions. For patients with cognitive and motor impairments, physicians may focus on their specialized rehabilitation needs. The referral decisions of physicians may be influenced by the need for and availability of specialized services. In addition, when dealing with patients with multiple functional impairments, physicians tend to refer them to higher-level medical institutions to access more comprehensive and complex rehabilitation services [21]. These differentiated referral decisions reflect physicians' consideration of the type of functional impairment in providing more specialized, personalized, and comprehensive rehabilitation services [22]. This difference may be influenced by specialization, resource allocation, and the complexity of multiple functional impairments.

Although this study primarily used multinomial logistic regression to identify independent predictors of referral destination, we recognize that an essential aim of a stratified care model is to ensure that case complexity appropriately drives referrals to higher-tier institutions. To explore this alignment, we performed a post hoc stratification analysis that compared the referral destinations among patients categorized by complexity indicators—such as the presence of multiple functional impairments, poor disease control, or bedridden status. The results showed that a substantial proportion of patients with high complexity (eg, Longshi bedridden group or unstable vital signs) were indeed referred to tertiary institutions, suggesting general consistency with stratified care principles. However, we also observed that certain high-need groups—such as patients with cognitive impairments—were frequently referred to secondary institutions. This raises the possibility of partial misalignment between patient complexity and referral level, potentially due to institutional resource constraints or physician referral preferences. While regression modeling helps quantify associations, future studies should consider incorporating appropriateness analyses or concordance rates to directly evaluate the extent to which physician decisions adhere to stratified care principles. This would offer a more practice-oriented perspective and strengthen evidence for policy-level recommendations.

The findings of this study are generally consistent with previous literature that highlights the influence of disease type and functional impairment on referral decisions. For example, Longley et al [20] emphasized that patients with severe mobility limitations were more likely to be directed to tertiary rehabilitation centers, aligning with our results regarding bedridden and domestically dependent patients. Similarly, Guo [17] reported that physicians often consider disease controllability and chronicity when determining the referral pathway, supporting our observation that patients with poor disease control were more often triaged to higher-level institutions.

However, some discrepancies exist. For instance, Labberton et al [21] found that cognitive impairments were strongly associated with referrals to tertiary care, whereas our study showed that such patients were frequently referred to secondary institutions. This divergence may stem from institutional differences in service structure—where secondary hospitals in China increasingly offer specialized neurocognitive rehabilitation services, reducing the need for tertiary-level referral. In addition, previous studies in developed healthcare systems such as Norway and Canada often emphasize patient autonomy and shared decision-making in the referral process, which contrasts with the physician-centered model observed in China. These differences underscore the importance of contextual factors—such as institutional capacity, policy incentives, and clinical practice norms—in shaping referral behavior.

The results of the Longshi Scale assessment indicated that patients assessed as bedridden or homebound were more likely to be referred to tertiary medical institutions. This may be because such assessments indicate significant functional limitations in daily activities [23]. These limitations might require more specialized and detailed rehabilitation services, leading physicians to refer these patients to higher-level institutions capable of providing more comprehensive care. The Longshi Scale assessment results provide physicians with a strong basis for understanding patients' daily living abilities and rehabilitation needs, aiding in more precise and comprehensive rehabilitation treatment decisions.

In addition to clinical and functional characteristics of patients, physicians' demographic and academic attributes may also play a significant role in shaping referral decision-making. Although this study collected comprehensive data on physicians' gender, age, education level, professional title, years of experience, and medical specialty, these variables were not included in the primary statistical analysis. This represents a limitation of the current study, as such factors could offer important perspectives on potential variability in referral behaviors across physician subgroups. For example, senior physicians with more than 20 years of clinical experience may demonstrate different triage patterns compared to entry-level physicians, possibly due to differences in clinical judgment, familiarity with tiered systems,

XSL•FO
RenderX

or risk tolerance. Similarly, physicians trained in traditional Chinese medicine or integrative medicine may prioritize different rehabilitation needs or institutional choices than their counterparts trained in Western medicine. Future research should incorporate multivariable analyses to explore how these practitioner characteristics intersect with referral decisions. Such analyses could uncover underlying patterns that inform more targeted physician training and policy interventions aimed at standardizing referral practices across diverse provider profiles.

To address whether more complex cases are in fact being referred to tertiary care facilities—a concern noted in recent literature and clinical guidelines—we conducted a focused review of the referral patterns based on proxies for case complexity. These included clinical indicators such as multiple functional impairments, uncontrolled disease status, unstable vital signs, and low Longshi Scale scores (eg, bedridden status). Our data demonstrated that patients exhibiting these complex characteristics were more likely to be referred to tertiary institutions. For example, patients classified as bedridden on the Longshi Scale were significantly more likely to be triaged to tertiary-level care (OR 0.10, 95% CI 0.14 - 0.34; $P<.001$), reflecting a recognition of their higher rehabilitation needs. Similarly, patients with uncontrolled disease or unstable vital signs were disproportionately represented among tertiary referrals, suggesting that physicians do consider clinical severity and complexity in their referral decisions. These findings support the appropriateness of current referral practices to some extent. However, there remains a subset of patients with cognitive and multiple impairments who were frequently referred to secondary rather than tertiary institutions, raising questions about possible mismatches between patient needs and referral destinations. This highlights the necessity for clearer referral criteria and physician decision-support tools to improve referral consistency and precision.

Recommendations

Enhancing Physician Training and Education

To improve physicians' decision-making in rehabilitation tiered medical services, it is recommended to strengthen their training and education [24]. This involves regular specialized lectures, seminars, and online learning platforms to provide physicians with the latest clinical guidelines and research findings, particularly focusing on key factors such as multiple functional impairments, disease control, vital signs, and Longshi Scale assessment. By continuously improving physicians' professional levels, their decision-making accuracy and scientific basis in rehabilitation tiered medical services can be enhanced.

Establishing Clear and Scientific Tiered Standards

It is suggested to develop more explicit and scientific tiered standards within the rehabilitation tiered medical service system. This requires in-depth research on patients' rehabilitation needs, revising promotable assessment tools and standards, and incorporating expert consensus. The standard-setting process should track the latest research findings in rehabilitation, address the diversity of patient needs, and consider changes in patients' rehabilitation needs at different stages. Clear referral tools for rehabilitation tiered medical services help physicians make more accurate referral decisions, increasing the scientific and precise nature of the system and better meeting the rehabilitation needs of patients in various medical institutions [25].

Promoting Information Technology Application

To enhance the scientific basis and data support for physicians' decisions in tiered medical services, the active promotion of information technology is recommended. Introducing electronic health record (EHR) systems to achieve comprehensive and accurate management of patient medical information [26]. Through digitalization, physicians can access key patient information more promptly, improving decision-making accuracy. Integrating assessment tools and standards related to rehabilitation tiered medical services into EHR systems provides physicians with more comprehensive, multidimensional analytical bases [27]. The application of information technology is expected to enhance the scientific and accurate nature of physicians' referral decisions while promoting the sharing and collaboration of medical information, thus improving the efficiency and quality of the rehabilitation tiered medical service system.

Limitations

This study focuses on the decision-making of rehabilitation outpatient physicians in China regarding patient referrals and analyzes the factors influencing their decisions, partially filling the research gap on outpatient referral decisions by rehabilitation physicians. This study is innovative to some extent. However, it still has certain limitations. First, this study used convenience sampling and included only 12 medical institutions in 5 cities, which might introduce selection bias and limit the generalizability of the results. Second, most of the outpatient patients included in this study were from the general population and primarily had motor function impairments, with relatively mild cognitive, speech, cardiopulmonary, and swallowing impairments. Therefore, the representativeness of the study sample is limited. Third, this study only used structured questionnaires for data collection, with a limited range of variables. Future research should consider conducting qualitative interviews with physicians to deeply explore the underlying factors influencing their patient referral decisions. Finally, although systemic factors (eg, health care policies and reimbursement schemes) may shape physician decision-making, these variables were not formally assessed in this study. Future research should incorporate such contextual factors to better understand referral patterns within China's evolving rehabilitation landscape. A notable limitation of this study is that the majority of patient-related variables—such as disease status, functional impairment classification, and Longshi Scale results—were completed by physicians rather than the patients themselves. While this approach ensured consistency and clinical accuracy in classification, it limited the incorporation of the patient perspective into the analysis. Factors such as patients' personal preferences for care setting, perceived accessibility, cost considerations, and trust in specific institutions were not assessed. These elements may play a significant role in shaping actual referral outcomes and deserve further investigation. Future studies could benefit from integrating patient-reported data alongside clinical assessments to provide

XSL•FO
RenderX

a more comprehensive picture of referral decision-making. Mixed-methods approaches, including structured patient surveys or qualitative interviews, may help uncover how patient preferences align—or conflict—with physician recommendations within the context of a tiered rehabilitation system.

To enhance the generalizability of future research, it is recommended that studies incorporate more diverse patient populations across broader geographic regions, including rural and under-resourced areas. This would help account for potential variations in referral practices that stem from regional disparities in medical resource availability, institutional capacity, and health policy implementation. In addition, stratified sampling strategies could be refined to include private healthcare institutions and specialized rehabilitation centers to better capture the full spectrum of referral behaviors.

Furthermore, while this study is situated within the context of China's public-tiered health care system, its findings may offer insights applicable to other countries pursuing similar rehabilitation reforms. For instance, countries with decentralized or mixed health care financing models—such as Brazil, India, or South Africa—face analogous challenges in aligning patient needs with institutional capacity. The clinical and functional indicators identified in this study (eg, disease control, Longshi Scale classification, and multi-impairment status) may serve as a reference framework for developing locally adapted referral tools. Future comparative studies could examine how such models perform across different health systems, ultimately contributing to a more globally informed understanding of rehabilitation referral strategies.

To operationalize these recommendations, several implementation strategies are proposed. First, integrating decision-support algorithms into existing EHR systems can help standardize physician referral behavior. For example, by embedding Longshi Scale results and disease-specific severity indicators into EHR templates, the system can automatically suggest referral destinations based on predefined criteria, thereby enhancing consistency and reducing subjectivity in triage decisions. This function could be further enhanced with real-time alerts or dashboard-style interfaces that highlight mismatches between patient status and current institutional capacity.

Second, the application of information technology should be piloted through regional health platforms that link primary, secondary, and tertiary institutions. For instance, the "Hierarchical Diagnosis and Treatment Information Platform" successfully deployed in cities like Shanghai and Chengdu has enabled real-time patient transfer, data sharing, and insurance coordination across institutional levels. These models can serve as scalable blueprints for implementing rehabilitation referral systems nationwide.

Finally, developing regional rehabilitation referral centers that provide training, feedback loops, and case review systems may enhance physician awareness and adherence to tiered service pathways. Such mechanisms, combined with policy incentives and reimbursement alignment, could significantly improve the feasibility and sustainability of the proposed tiered rehabilitation model in China's evolving health care landscape.

Conclusions

In the field of rehabilitation, implementing tiered medical services is crucial for the rehabilitation outcomes of patients and the rational use of medical resources. Rehabilitation tiered medical services can effectively triage patients based on comprehensive, accurate, and standardized referral decision tools. This helps ensure that patients receive more precise and professional rehabilitation treatment at suitable medical institutions, improving treatment outcomes and enhancing the convenience of seeking medical care for patients. Furthermore, this practice aids the government in rationalizing the use of medical resources, improving the efficiency and quality of medical services, and reducing health care costs. To thoroughly and comprehensively understand the key factors influencing physicians' decisions to refer rehabilitation patients, and to develop standardized rehabilitation tiered medical service tools, future research could consider conducting large-scale, multicenter, cluster-randomized controlled trials at the city and national levels. This would facilitate better optimization and development of the rehabilitation tiered medical service system.

Despite its strengths, this study has certain limitations that may affect the generalizability of its findings. The use of convenience sampling and inclusion of institutions from only 5 cities may introduce selection bias. In addition, the analysis did not include system-level variables such as policy implementation strength or institutional resource availability. These factors should be examined in future studies to enhance the comprehensiveness of the referral decision framework. Nonetheless, this study makes a valuable contribution by being one of the first large-scale, multicenter investigations into the determinants of physician referral decisions within China's rehabilitation tiered health care system. By identifying specific clinical and functional characteristics associated with referral destinations, it provides an empirical foundation for developing standardized referral tools, refining triage protocols, and informing policy design in the field of rehabilitation medicine.

Acknowledgments

We are grateful to the individuals from each medical institution for their support and advice in the provision of data. This research was funded by the Sanming Project of Medicine in Shenzhen (SZSM202111010), Shenzhen Key Medical Discipline Construction Fund (SZXK048), Shenzhen Science and Technology Program (JCYJ20240813141037048), and Shenzhen Health Economics Society 2025 Scientific Research Fund Project (2025343).

XSL•FO
RenderX

Data Availability

The datasets generated and analyzed during this study are not publicly available due to ethical restrictions imposed by participant consent agreements, but are available from the corresponding author upon reasonable request. All data sharing requests will be subject to review and approval by the Second People's Hospital of Shenzhen Ethical Review Board to ensure compliance with ethical protocols.

Authors' Contributions

YL contributed to conceptualization, methodology, software, formal analysis, writing the original draft, preparation, and investigation. RY contributed to validation, visualization, formal analysis, investigation, and writing – review and editing. LS and KW contributed to data curation and investigation. KX and ZZ performed formal analysis and investigation. YH contributed to writing – review and editing. YZ and XL contributed to writing – review and editing and validation. YW contributed to supervision, validation, project administration, resources, and funding acquisition.

Conflicts of Interest

None declared.

References

1. NHS Improvement annual report and accounts 2019/20. : NHS; 2020 URL: https://www.england.nhs.uk/publication/nhs-improvement-annual-report-and-accounts-2019-20/ [accessed 2024-06-04]
2. Australia's long term national health plan. Australian Government Department of Health. 2018. URL: https://www.health.gov.au/resources/publications/australias-long-term-national-health-plan [accessed 2024-06-04]
3. Health care system. Health Canada. 2021. URL: https://www.canada.ca/en/health-canada/services/health-care-system.html [accessed 2024-06-04]
4. Quality matters: realizing excellent care for all. Health Quality Ontario. 2019. URL: https://www.hqontario.ca/portals/0/documents/health-quality/realizing-excellent-care-for-all-en.pdf [accessed 2024-06-04]
5. Campbell SM, Roland MO, Buetow SA. Defining quality of care. Soc Sci Med 2000 Dec;51(11):1611-1625. [doi: 10.1016/S0277-9536(00)00057-5] [Medline: 30476741]
6. Smith SM, Wallace E, O'Dowd T, Fortin M. Interventions for improving outcomes in patients with multimorbidity in primary care and community settings. Cochrane Database Syst Rev 2019;1(1):CD006560. [doi: 10.1002/14651858.CD006560.pub4]
7. Blank L, Baxter S, Woods HB, et al. Referral interventions from primary to specialist care: a systematic review of international evidence. Br J Gen Pract 2014 Dec;64(629):e765-e774. [doi: 10.3399/bjgp14X682837] [Medline: 25452541]
8. Guiding opinions of the general office of the state council on promoting the construction of the hierarchical medical system [Web page in Chinese]. State Council of China. 2015. URL: http://www.gov.cn/zhengce/content/2015-09/11/content_10158.htm [accessed 2024-06-04]
9. Wang C, Li J, Zhao X, Wang Y, Wu D, Wang Y. Stroke care development in China: past, present and future. Int J Stroke 2008 Nov;3(4):288-289. [doi: 10.1111/j.1747-4949.2008.00218.x]
10. Yan N, Liu T, Xu Y, et al. Healthcare preferences of the general Chinese population in the hierarchical medical system: a discrete choice experiment. Front Public Health 2022;10:1044550. [doi: 10.3389/fpubh.2022.1044550] [Medline: 36466449]
11. McGinn TG, McCullagh L, Kannry J, et al. Efficacy of an evidence-based clinical decision support in primary care practices: a randomized clinical trial. JAMA Intern Med 2013 Sep 23;173(17):1584-1591. [doi: 10.1001/jamainternmed.2013.8980] [Medline: 23896675]
12. Zhan YL, Li YY, Li P, et al. Analysis of the current status and efficacy of acute ischemic stroke treatment based on regional hierarchical medical service network. Journal of Capital Medical University 2025. [doi: 10.3969/j.issn.1006-7795.2025.01.007]
13. Wang D, Li HW, Yuan BB, et al. Analysis on the primary care doctor's referral behavior and its influencing factors in urbanand rural areas of China. Chinese Journal of Health Policy 2019;12(9):25-30.
14. Zhang YL, Wang L, Xu C, et al. Analysis of influencing factors for the implementation of "dual diagnosis system" in Chaoyang District. Beijing [J] Chinese General Practice 2010;13(31):3507-3509. [doi: 10.3969/j.issn.1007-9572.2010.31.011]
15. Feng N. Investigation on knowledge, attitude, and practice of hierarchical medical system among doctors at all levels of hospitals in xinjiang and research on improvement strategies. Xinjiang Medical University 2019. [doi: 10.3969/ji.sn.1009-5551.2019.03.026]
16. Wang YL. Application of intelligent disability rating method based on big data cloud platform. RM 2022 Aug 1;32(4):285-292. [doi: 10.3724/SP.J.1329.2022.04001]
17. Guo YP. Research on hierarchical healthcare based on patients' healthcare choice [Dissertation in Chinese]. : Yanshan University; 2019 URL: https://tinyurl.com/5ensm8r7 [accessed 2024-06-04]
18. Tang C. Challenges and suggestions for orthopedic doctors in grassroots hospitals. Chinese Journal of Traditional Chinese Medicine Management 2016;24(12):106-107. [doi: 10.16690/j.cnki.1007-9203.2016.12.053]

XSL•FO
RenderX

19. Li YN, Wang YQ. Prevalence and patterns of multimorbidity among Chinese elderly people. Chinese General Practice 2021;24(31):3955-3962. [doi: 10.12114/j.issn.1007-9572.2021.00.295]
20. Longley V, Peters S, Swarbrick C, Bowen A. What factors affect clinical decision-making about access to stroke rehabilitation? A systematic review. Clin Rehabil 2019 Feb;33(2):304-316. [doi: 10.1177/0269215518808000] [Medline: 30370792]
21. Labberton AS, Barra M, Rønning OM, et al. Patient and service factors associated with referral and admission to inpatient rehabilitation after the acute phase of stroke in Australia and Norway. BMC Health Serv Res 2019 Nov 21;19(1):871. [doi: 10.1186/s12913-019-4713-x] [Medline: 31752874]
22. Wang ZJ, Yu HJ, Tang J, et al. Cognitive impairment rate of the elderly in China: a meta-analysis [Article in Chinese]. Chinese Journal of Evidence-Based Medicine 2020;20(11):1295-1300 [FREE Full text] [doi: 10.7507/1672-2531.202006099]
23. Lv HL, Li F, Wang Y, et al. Exploring the care needs of long-term bedridden patients in China based on the Long's scale [Article in Chinese]. Chinese Medicine and Pharmacy 2023;13(12):87-91 [FREE Full text]
24. Li WM, Li WZ, Zeng XY. Research on the key issues affecting the promotion of hierarchical medical system in China. Chinese Health Policy Research 2018;11(10):12-19. [doi: 10.3969/j.issn.1674-2982.2018.10.002]
25. Li ZR, Wei RM, Xu Y, et al. The exploration of hierarchical medical system from the perspective of supply side. Chinese Health Economics 2018;37(2):14-17. [doi: 10.7664/CHE20180204]
26. Chen XH, Li J, Cui J, et al. Operational mechanism of tele- hierarchical care：analysis from the complex adaptive system [Article in Chinese]. Chinese Hospital Management 2023;43(7):20-24 [FREE Full text]
27. Pirruccello JP, Traynor KC, Natarajan P, et al. An electronic cardiac rehabilitation referral system increases cardiac rehabilitation referrals. Coron Artery Dis 2017 Jun;28(4):342-345. [doi: 10.1097/MCA.0000000000000491] [Medline: 28346284]

Abbreviations

EHR: electronic health record
OR: odds ratio
VIF: variance inflation factor

Edited by TDA Cardoso; submitted 18.07.24; peer-reviewed by K Ribbons, SU Uzun; revised version received 23.05.25; accepted 26.05.25; published 24.07.25.

Please cite as:
Li Y, Ye R, Shan L, Wang K, Xue K, Zhang Z, Hao Y, Zou Y, Li X, Wang Y
Factors Influencing Physicians' Referral Decision-Making for Rehabilitation Outpatient Services in the Health Care Landscape of China: Cross-Sectional Study
Interact J Med Res 2025;14:e64464
URL: https://www.i-jmr.org/2025/1/e64464
doi:10.2196/64464

© Yawei Li, Ruixue Ye, Linlin Shan, Kun Wang, Kaiwen Xue, Zeyu Zhang, Yingzi Hao, Yucong Zou, Xiaoxuan Li, Yulong Wang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 24.7.2025. 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

Policy Spotlight Effects on Critical Time-Sensitive Diseases: Nationwide Retrospective Cohort Study on Taiwan's Hospital Emergency Capability Categorization Policy

Chih-Yuan Lin[1,2], MSc, MD, PhD; Chih-Ching Liu[3], PhD; Yu-Tung Huang[4], PhD; Yue-Chune Lee[2], DrPH

[1]Department of Neurology, Taipei City Hospital Linsen Chinese Medicine Branch, Taipei City, Taiwan

[2]Institute of Health and Welfare, College of Medicine, Yangming Campus, National Yang Ming Chiao Tung University, Taipei, Taiwan, Taipei City, Taiwan

[3]Department of Health Care Management, National Taipei University of Nursing and Health Science, Taipei, Taiwan

[4]National Center for Geriatrics and Welfare Research, National Health Research Institutes, Yunlin, Taiwan

Corresponding Author:
Yue-Chune Lee, DrPH
Institute of Health and Welfare, College of Medicine, Yangming Campus
National Yang Ming Chiao Tung University, Taipei, Taiwan
No.155, Sec.2, Linong Street
Taipei City, 112
Taiwan
Phone: 886 225916681 ext 1526
Fax: 886 225916681
Email: DAS73@tpech.gov.tw

Abstract

Background: Taiwan's categorization of hospital emergency capability (CHEC) policy is designed to regionalize and dispatch critical patients. The policy was designed in 2009 to improve the quality of emergency care for critical time-sensitive diseases (CTSDs). The CHEC policy primarily uses time-based quality surveillance indicators.

Objective: We aimed to investigate the impact of Taiwan's CHEC policy on CTSDs.

Methods: Using Taiwan's 2005 Longitudinal Health Insurance Database, this nationwide retrospective cohort study examined the CHEC policy's impact from 2005 to 2011. Propensity score matching and difference-in-differences analysis within a generalized estimating equation framework were used to compare pre- and postimplementation periods. The study focused on acute ischemic stroke (AIS), ST-segment elevation myocardial infarction (STEMI), septic shock, and major trauma. AIS and STEMI cases, monitored with time-based indicators, were evaluated for adherence to diagnostic and treatment guidelines as process quality measures. Mortality and medical use served as outcome indicators. Major trauma, with evolving guidelines and no time-based monitoring, acted as a control to test for policy spotlight effects.

Results: In our cohort of 9923 patients, refined through 1:1 propensity score matching, 5566 (56.09%) were male and were mostly older adults. Our analysis revealed that the CHEC policy effectively improved system efficiency and patient outcomes, resulting in significant reductions in medical orders (–7.29 items, 95% CI –10.09 to –4.48; *P*<.001), short-term mortality rates (–0.09%, 95% CI –0.17% to –0.02%; *P*=.01) and long-term mortality rates (–0.09%, 95% CI –0.15% to –0.04%; *P*=.001), and total medical expenses (–5328.35 points per case, 95% CI –10,387.10 to –269.60; *P*=.04), despite a modest increase in diagnostic fees (376.37 points, 95% CI 92.42-660.33; *P*=.01). The CHEC policy led to notable increases in diagnostic fees, major treatments, and medical orders for AIS and STEMI cases. For AIS cases, significant increases were observed in major treatments (β=0.77; 95% CI 0.21-1.33; *P*=.007) and medical orders (β=15.20; 95% CI 5.28-25.11; *P*=.003) compared to major trauma. In STEMI cases, diagnostic fees significantly increased (β=1983.75; 95% CI 84.28-3883.21; *P*=.04), while upward transfer rates significantly decreased (β=–0.59; 95% CI –1.18 to –0.001; *P*=.049). There were also trends toward increased major treatments (β=0.30; 95% CI –0.03 to 0.62, *P*=.07), medical orders (β=11.92; 95% CI –0.90 to 24.73; *P*=.07), and medical expenses (β=24,275.54; 95% CI –640.71 to 4,991,991.78; *P*=.06), although these were not statistically significant. In contrast, no significant changes were identified in process or outcome quality indicators for septic shock. These findings suggest policy spotlight effects, reflecting a greater emphasis on diseases directly prioritized under the CHEC policy.

XSL•FO
RenderX

Conclusions: The CHEC policy demonstrated the dual benefits of reducing costs and improving patient outcomes. We observed unintended consequences of policy spotlight effects, which led to a disproportionate improvement in guideline adherence and process quality for CTSDs with time-based surveillance indicators.

(Interact J Med Res 2025;14:e54651) doi:10.2196/54651

KEYWORDS

categorization of hospital emergency capability; quality; time-sensitive diseases; emergency care; difference-in-differences

Introduction

Background

The Agency for Healthcare Research and Quality proposed the concept of time-sensitive diseases [1] using scientific data to maintain up-to-date guidelines and launched the Get with the Guidelines campaign to establish it as the basis for surveillance indicators of process and outcome quality [2]. Critical time-sensitive diseases (CTSDs) refer to life-threatening illnesses or injuries that require immediate emergency care, where rapid intervention is paramount to mitigate morbidity and mortality [3]. The various guidelines for managing time-sensitive events emphasize the crucial importance of time. In the context of acute ischemic stroke (AIS), the "time is brain" [4] goal focuses on the timely reperfusion treatments, including intravenous thrombolysis and mechanical thrombectomy; in ST-segment elevation myocardial infarction (STEMI), the "time is muscle" goal focuses on early reperfusion [5]; in septic shock events, the "early goal" focuses on early resuscitation [6]; and in major trauma cases, the "golden hour" goal focuses on the window of opportunity in which patients can undergo rescue operations [7].

The American Medical Association issued the categorization of hospital emergency capability (CHEC) guidelines [8] to classify hospitals according to their emergency care capabilities, thereby regionalizing and providing emergency medical services with references for transport emergency patients to the nearest appropriate hospitals [9], aiming to reduce preventable deaths. Most studies investigating the effects of this categorization, designation, and regionalization policy reported positive findings [10,11]. However, these studies mainly focused on a single disease entity [10,11] or region [11]. The CHEC policy often implements rigid time-based surveillance indicators. These indicators can affect disease-specific guideline adherence in clinical practice because they may reshape the behaviors of emergency department (ED) medical providers [12]. This phenomenon is related to the so-called policy spotlight effects, which influence medical care providers' assessment of how others perceive them [13]. More specifically, the policy spotlight effects refer to the perception of medical care providers regarding how policy makers interpret surveillance indicators and adjust their process-related behaviors accordingly [14]. Current emergency care policies often use time-based criteria as process quality indicators, which may exacerbate the policy spotlight effects [13]; however, the unintended effects or safety concerns generated by these effects remain unclear. Therefore, our study targeted 4 CTSDs: AIS, STEMI, septic shock, and major trauma [7]. Our research hypothesizes that emergency care providers might inadvertently give more attention to diseases under active surveillance while potentially neglecting those not thoroughly incorporated in this observation. This focus might be based on their perception of observer expectations [15].

Objective

We aimed to examine the effects of the CHEC policy on process quality and outcomes for CTSDs, addressing three research questions: (1) How does the CHEC policy impact the quality of diagnosis, treatment, and outcomes for these diseases? (2) How do policy spotlight effects influence the prioritization of diseases under active surveillance and impact emergency care providers' behaviors in this context? and (3) What are the potential consequences of policy spotlight effects?

Methods

Setting, Study Design, and Data Source

Taiwan's National Health Insurance is a single-payer, compulsory social insurance system that primarily operates on a fee-for-service basis. This study is based on the National Health Insurance 2005 Longitudinal Health Insurance Database (LHID2005), which contains 1 million random cases, including medical records and hospital information, collected since 1995. The LHID2005 was validated to represent medical use, diagnosis and treatment process, and outcome quality for CTSDs [16].

This nationwide retrospective cohort study uses propensity score matching (PSM) and difference-in-differences (DID) analysis to evaluate the impact of the CHEC policy on CTSD care quality and outcomes. The CHEC policy was initiated in August 2009, which integrated 190 hospitals into a network focusing on acute conditions such as stroke, myocardial infarction, major trauma, and perinatal care [17]. We divided our analysis into 2 periods: before CHEC (August 1, 2005, to July 31, 2009) and after CHEC (August 1, 2009, to July 31, 2011). This division aimed to assess the CHEC policy's effects distinctly from the ED quality improvement plan introduced in 2012. Well-established guidelines exist for AIS, STEMI, and septic shock. In contrast, the guidelines for major trauma are continuously evolving due to the variability in injury mechanisms, locations, and severity. Moreover, AIS and STEMI events are stringently monitored under the CHEC policy with specific time-based quality indicators, whereas septic shock and major trauma events are not (Table 1).

XSL•FO
RenderX

Table 1. Critical time-sensitive diseases and categorization hospital emergency capability (CHEC) policy indicators in Taiwan [18].

Quality indicator	Acute ischemic stroke	ST-segment elevation MI[a]	Septic shock	Major trauma
Guidelines development	• Well developed	• Well developed	• Well developed	• Developing
Major diagnosis indicator	• Brain imaging (eg, CT[b] and MRI[c])	• EKG[d]	• Blood culture	• Image study
Major treatment indicator	• Intravenous TPA[e]	• PCI[f]	• Antibiotics	• Rescue operation
Guideline's major goal	• Early thrombolysis • Time is brain	• Early reperfusion • Time is muscle	• Goal-directed therapy • Early goal	• Rescue operation • (Golden hour)
Guideline's time-based criteria	• 60 min	• 90 min	• 3 to 6 h	• 1 h
CHEC policy indicators	• Stroke team • NIHSS[g] score evaluation • Intravenous TPA	• PCI team • Give aspirin and clopidogrel	• ICUS[h] critical care team	• Trauma team • ISSi evaluation
CHEC policy time-based criteria	• Neurologist consultation in <30 min • Door to CT in <30 min • Door to CT read in <45 min • Onset to needle in <3 h	• Cardiologist consultation in <30 min • Door to EKG in <10 min • Door to needle in <30 min • Door to balloon in <90 min	• Admission in <24 h	• Trauma team activation in <30 min

[a]MI: myocardial infarction.

[b]CT: computed tomography.

[c]MRI: magnetic resonance imaging.

[d]EKG: electrocardiography.

[e]TPA: tissue plasminogen activator.

[f]PCI: percutaneous coronary intervention.

[g]NIHSS: National Institute of Health Stroke Scale or Score.

[h]ICU: intensive care unit.

[i]ISS: Injury Severity Score.

Thus, we selected major trauma events as a reference for our study because they were not monitored under the CHEC policy with rigid indicators. We adopted pre- and postimplementation of a CHEC policy, using 1:1 PSM to control for confounding variables. We used the DID estimation approach to estimate the association of the CHEC policy on process and outcomes for AIS and STEMI. For the counterfactual, we used major trauma cases unexposed to the clinical guideline or CHEC policy time-based quality indicators as the basis for comparison.

Identification of Study Cohort

This study identified CTSDs based on ED visits accompanied by a primary diagnosis using the appropriate disease code. The identification of AIS (ie, codes 433 and 434), STEMI (ie, code 410), and septic shock (ie, codes 038, 785, and 995) was based on the *International Classification of Diseases, Ninth Revision, Clinical Modification* (*ICD - 9 - CM*). Major trauma cases were classified following the American Academy of Surgery Committee guidelines (ie, codes 800-959) [19]. Due to the absence of trauma severity data in the LHID2005 database, primary *ICD-9-CM* codes served as our initial method for identifying major trauma incidents. This identification was further refined by including cases where patients received rescue surgery or were admitted to the intensive care unit, serving as additional criteria for major trauma [20]. We excluded cases before the study period and those without hospital or patient sociodemographic information. We also excluded hospitals with a volume of <5 CTSD cases per year [21]. We used the date of the first ED visit as the index date.

Definition of Variables

The independent variable in this study was exposure to the CHEC policy intervention. Events related to AIS and STEMI were subject to rigid time-based quality indicators and regular surveillance under the CHEC policy. In contrast, despite having well-developed clinical guidelines, septic shock was not included under the CHEC policy's stringent time-based quality indicators. Similarly, major trauma cases, which lack well-developed clinical guidelines, were not subject to these policy indicators and were used as counterfactuals in this study. The dependent variables were divided into primary and secondary outcomes. Primary outcomes included guideline adherence to diagnostic and treatment process quality indicators. For diagnostic adherence, AIS was assessed by the completion of brain imaging (eg, computed tomography or magnetic resonance imaging) within 60 minutes of hospital arrival, while

XSL•FO
RenderX

STEMI was evaluated based on the completion of electrocardiography within 10 minutes. For treatment adherence, AIS required the administration of intravenous thrombolysis within 3 hours of symptom onset, and STEMI was assessed according to whether percutaneous coronary intervention was performed within 90 minutes of hospital arrival. Secondary outcomes included upward transfer rates, diagnostic fees, medical orders and expenses, and mortality rates. Upward transfer rates were defined as the proportion of patients transferred from lower-level hospitals to higher-level facilities. Diagnostic fees referred to the total costs incurred from diagnostic procedures during ED care. Medical orders and expenses represented the number and costs of medical interventions performed during ED visits. Mortality rates were measured as 30-day mortality, indicating the percentage of patients who died within 30 days from the index date (the emergency visit date), and 1-year mortality, representing deaths occurring within 1 year.

Covariates included patient-related predisposing factors, such as age, sex, and occupation. The enabling factor was the insured salary, while the Charlson Comorbidity Index, calculated using *ICD-9-CM* codes from primary diagnoses recorded in inpatient and outpatient claims data up to a year before the index date, served as a measure of health needs. Furthermore, external environmental factors, including urbanization and regional emergency resources, were considered. For hospital-level variables, the input-throughput-output model of Asplin et al [22] was used. Input was gauged using the rate of ED visits with Emergency Severity Index 1, while throughput and output efficiency were assessed via the ED's occupancy rate. This comprehensive framework ensured a robust evaluation of the CHEC policy's impact while accounting for potential confounding variables and contextual factors.

Statistical Analysis

Patients' characteristics, process quality, and outcomes were presented using descriptive statistics. Continuous data were described using mean (SD), and categorical data were presented using numbers and percentages. To enhance the robustness of the outcomes, we calculated the propensity score using a multivariable logistic regression that included all baseline covariates. The standardized mean difference was calculated to confirm the balance of potential confounders at baseline between groups before and after matching. A standardized mean difference of <0.1 was considered to represent a negligible difference [23].

We evaluated the impact of the CHEC intervention on each outcome, including overall differences, the differences within individual diseases, and between-disease differences in changes from baseline (ie, group-by-disease interaction effects), using a DID framework integrated with generalized estimating equation (GEE) models. The DID approach allowed us to compare changes in outcomes between the before and after CHEC policy periods across diseases, enabling the estimation of differential effects of the intervention while controlling for time-invariant unobserved confounders. This method is particularly suitable for evaluating policy interventions by focusing on within-group changes relative to a reference group over time. The GEE model is specified as follows:

$$Y_{ij} = \beta_0 + \beta_1 (\text{CHEC policy})_j + \beta_2 \text{ disease } + \beta_3 (\text{CHEC policy} \times \text{disease})_{ij} + \varepsilon_{ij,}$$

where Y_{ij} represents the outcome variable (eg, diagnosis indicator, treatment indicator, mortality, or medical use) for individual i at time j. β_0 is the intercept, capturing the baseline level of the outcome for the reference disease (ie, major trauma) in the pre-CHEC period. CHEC policy$_j$ is a binary variable (0=before CHEC, 1=after CHEC), and its coefficient β_1 captures the overall impact of the CHEC policy across all diseases. Disease$_i$ is a categorical variable representing the 4 diseases (ie, AIS, STEMI, septic shock, and major trauma), with β_2 estimating disease-specific differences at baseline. The interaction term (CHEC policy × disease)$_{ij}$, with coefficient β_3, reflects the differential impact of the CHEC policy for each disease compared to the reference group (ie, major trauma). A positive β_3 value indicates a greater change in the outcome for the respective disease relative to major trauma. All statistical analyses were performed using SAS (version 9.4; SAS Institute), and statistical significance was defined as value of <.05.

Ethical Considerations

This study used secondary data from the NIH LHID2005, which are fully anonymized and deidentified to protect participant privacy. The dataset contained no personal identifiers, such as names, addresses, or social security numbers, so individual informed consent was not required. The original consent provided for primary data collection, and the institutional review board approval covered secondary analyses without requiring additional consent. All analyses were conducted in compliance with relevant regulations to safeguard confidentiality, and data access was restricted to authorized researchers under institutional guidelines. The study received ethics approval from the Taiwan National Yang-Ming University Institutional Review Board (YM107035E).

Results

Participants' Characteristics

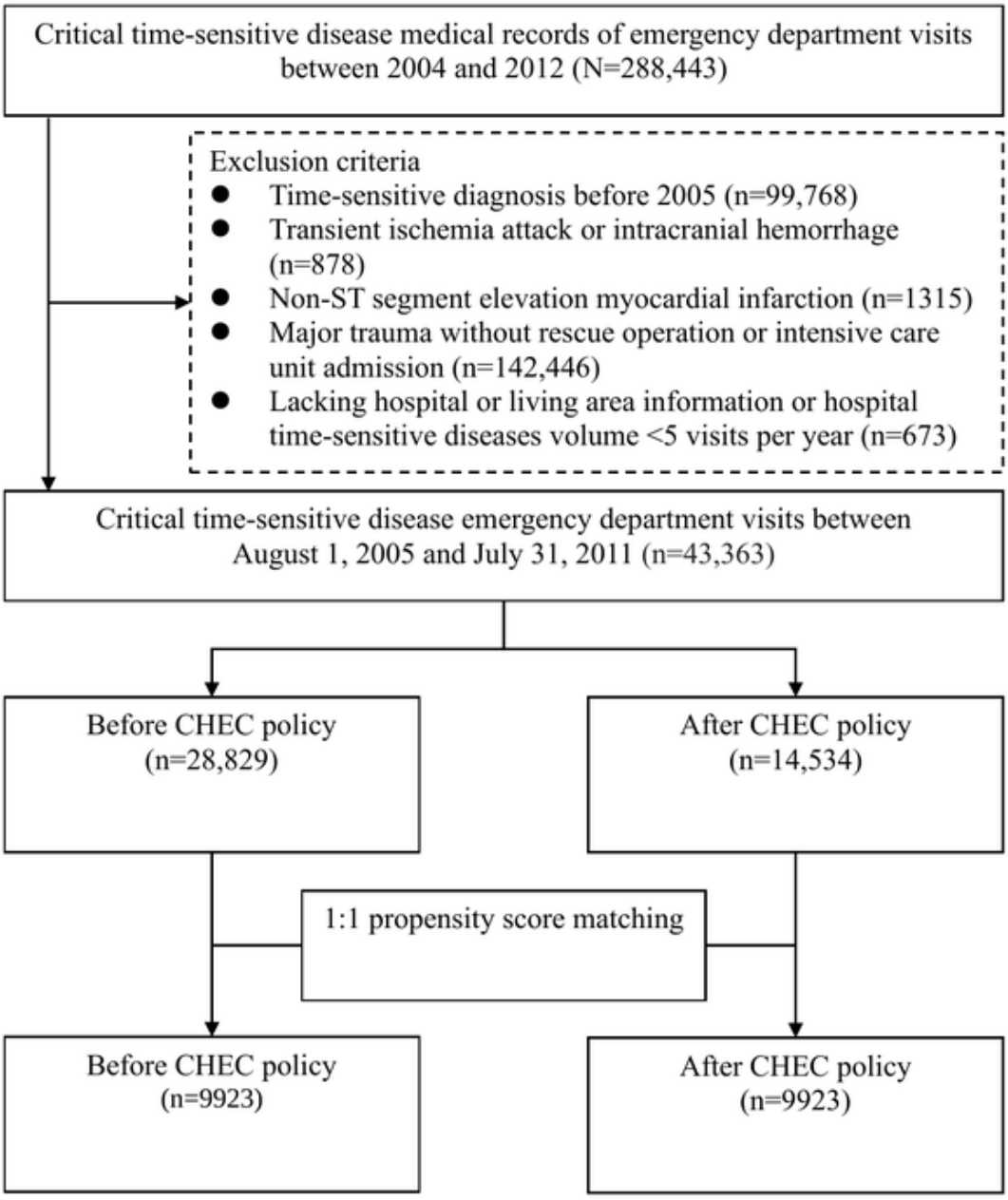
During the study period, we analyzed emergency presentations related to 4 CTSDs, originally encompassing 288,443 patients. Exclusion criteria included the diagnosis of CTSDs before 2005 (n=99,768, 34.59%), patients with transient ischemic attack or intracranial hemorrhage (n=878, 0.3%), non-STEMI cases (n=1315, 0.45%), individuals with major traumas defined by *ICD-9-CM* codes that did not necessitate a rescue operation or intensive care unit admission (n=142,446, 49.38%), and cases lacking hospital or living area information or where the hospital's volume of CTSDs was <5 visits per year (n=673, 0.23%). These criteria refined the total sample size to 43,363 (15.03%) patients. Considering the extended period before the policy intervention, this research adopted a 1:1 PSM technique, resulting in a final matched sample of 9923 patients. The flowchart and baseline table (Figure 1) display the initial count of emergency patients with CTSDs and the numbers after PSM, broken down by each of the 4 diseases. Table 2 presents the

XSL•FO
RenderX

PSM of participants with CTSDs before and after the PSM. After the matching process, each variable baseline characteristic demonstrated almost complete congruity. In addition, uniformity was achieved within each disease subgroup after matching (Table S1 in Multimedia Appendix 1). In 9923 patients, the distribution for each disease before and after PSM was as follows: AIS (n=2895, 29.17%), STEMI (n=723, 7.29%), septic shock (n=5441, 54.83%), and major trauma (n=864, 8.71%).

Septic shock was the most prevalent condition, accounting for 54.83% (5441/9923) of all cases. The patient population was male-dominated (5566/9923, 56.09%), with the majority (6084/9923, 61.31%) aged ≥65 years. Nearly three-fourths of the CTSD cases (7563/9923, 76.12%) were managed in hospitals categorized as moderate or severe levels. Care provided by specialty consultants accounted for 67.9% (6738/9923) of the cases.

Figure 1. Flow diagram of the study.



XSL•FO
RenderX

Table 2. Propensity score matched the comparison of patient characteristics before and after the categorization of hospital emergency capability (CHEC) policy. The complete table is available in Supplementary Table 2: Association of CHEC policy with process and outcome quality in four critical time-sensitive diseases (Multimedia Appendix 2).

Patient characteristics	Before propensity score matching			After propensity score matching		
	Before CHEC policy (n=28,829)	After CHEC policy (n=14,534)	ASMD[a]	Before CHEC policy (n=9923)	After CHEC policy (n=9923)	ASMD
Sex, n (%)						
Female	12,345 (42.82)	6250 (43)	0.004	4357 (43.91)	4268 (43.01)	0.018
Male	16,484 (57.18)	8284 (57)	0.004	5566 (56.09)	5655 (56.99)	0.018
Age (y), n (%)						
≤45	4105 (14.24)	1967 (13.53)	0.021	1289 (12.99)	1306 (13.16)	0.005
45-64	7173 (24.88)	3795 (26.11)	0.028	2550 (25.7)	2590 (26.1)	0.009
≥65	17,551 (60.88)	8772 (60.36)	0.011	6084 (61.31)	6027 (60.74)	0.012
Charlson Comorbidity Index, n (%)						
≤1	13,215 (45.84)	6546 (45.04)	0.016	4473 (45.08)	4434 (44.68)	0.008
>1	15,614 (54.16)	7988 (54.96)	0.016	5450 (54.92)	5489 (55.32)	0.008
Income[b] (NT $), n (%)						
≤22,800	15,961 (55.36)	6203 (42.68)	0.256	4428 (44.62)	4604 (46.4)	0.036
>22,800	12,868 (44.64)	8331 (57.32)	0.256	5495 (55.38)	5319 (53.6)	0.036
Occupation, n (%)						
Dependents of the insured	10,095 (35.02)	4952 (34.07)	0.020	3550 (35.78)	3326 (33.52)	0.048
Civil servants, teachers, and military personnel	2792 (9.68)	1529 (10.52)	0.028	986 (9.94)	1053 (10.61)	0.022
Nonmanual workers and professionals	2231 (7.74)	1240 (8.53)	0.029	748 (7.54)	796 (8.02)	0.018
Manual workers	10,250 (35.55)	5154 (35.46)	0.002	3422 (34.49)	3708 (37.37)	0.060
Others	3461 (12.01)	1659 (11.41)	0.019	1217 (12.26)	1040 (10.48)	0.056
Hospital categorization, n (%)						
Severe level	11,371 (39.44)	5924 (40.76)	0.027	3331 (33.57)	3502 (35.29)	0.036
Moderate level	10,921 (37.88)	5496 (37.81)	0.001	4030 (40.61)	4052 (40.83)	0.004
General level	6537 (22.68)	3114 (21.43)	0.030	2562 (25.82)	2369 (23.87)	0.045
ESI[c] triage level 1 and 2 rate, mean (SD)	5.8 (8.98)	16.63 (10.06)	1.136	14.01 (10.39)	14.37 (10.50)	0.034
Length of ED[d] stay, mean (SD)	5.33 (10.11)	4.49 (9.30)	0.086	6.25 (9.30)	5.36 (10.26)	0.091
ED observation ≥1-d rate, mean (SD)	11.37 (9.97)	12.32 (10.06)	0.095	10.53 (9.34)	11.32 (10.12)	0.081
Place of ED resources, n (%)						
Sufficiency	22,770 (78.98)	11,460 (78.85)	0.003	7768 (78.28)	7834 (78.95)	0.016
Not sufficiency	6059 (21.02)	3074 (21.15)	0.003	2155 (21.72)	2089 (21.05)	0.016
Time-sensitive disease, n (%)						
Acute ischemic stroke	8660 (30.04)	3814 (26.24)	0.085	2895 (29.17)	2895 (29.17)	0.000
ST-segment elevation MI[e]	2481 (8.61)	1141 (7.85)	0.028	723 (7.29)	723 (7.29)	0.000
Septic shock	14,896 (51.67)	8275 (56.94)	0.106	5441 (54.83)	5441 (54.83)	0.000
Major trauma	2792 (9.68)	1304 (8.97)	0.024	864 (8.71)	864 (8.71)	0.000

XSL•FO
RenderX

Patient characteristics	Before propensity score matching			After propensity score matching		
	Before CHEC policy (n=28,829)	After CHEC policy (n=14,534)	ASMD[a]	Before CHEC policy (n=9923)	After CHEC policy (n=9923)	ASMD
Delivery of care, n (%)						
Specialty consultant[f]	18,692 (64.84)	9245 (63.61)	0.026	6738 (67.9)	6570 (66.21)	0.036
Emergency physician	7788 (27.01)	4236 (29.15)	0.048	2449 (24.68)	2585 (26.05)	0.031
Others	2349 (8.15)	1053 (7.25)	0.034	736 (7.42)	768 (7.74)	0.012

[a]ASMD: absolute standardized mean difference.

[b]INT $1=US $0.03057 as of February 16, 2025.

[c]ESI: Emergency Severity Index.

[d]ED: emergency department.

[e]MI: myocardial infarction.

[f]Specialty consultant: (1) acute ischemic stroke is treated by neurologists, (2) acute MI is treated by cardiologists, (3) septic shock is managed by internal medicine physicians or critical care intensivists, and (4) major trauma conditions are managed by surgeons or critical care intensivists.

Impact of CHEC Policy Overall and on the Processes and Outcomes of the 4 Individual CTSDs Before and After Implementation

In examining individual diseases, primary diagnostic indicators for AIS, septic shock, and major trauma decreased after intervention, while only those for STEMI increased (Table 3).

Diagnostic fees increased for AIS, STEMI, and major trauma cases but decreased for septic shock cases. A similar trend was observed in the primary treatment indicators, which increased for AIS and STEMI cases and decreased for septic shock and major trauma cases. In contrast, medical orders showed a universal decline. Upward transfer rates increased for AIS and major trauma cases but decreased for STEMI and septic shock cases. Regarding outcome indicators, short-term and long-term mortality rates displayed a universal decline, except for AIS cases, which showed an increase. The medical expenses were higher for AIS and STEMI cases but lower for septic shock and major trauma cases.

In assessing the overall policy effects on 4 CTSD cohorts, the primary diagnosis indicator significantly decreased by 0.21% points (95% CI –0.29% to –0.13%; $P<.001$) and medical orders per case dropped by an average of 7.29 items (95% CI –10.09 to –4.48; $P<.001$). In comparison, diagnostic fees demonstrated an average increase of 376.37 points (95% CI 92.42-660.33; $P=.01$). The 30-day mortality rate saw a notable reduction of 0.09% points (95% CI –0.17% to –0.02%; $P=.01$), 1-year mortality significantly decreased by 0.09% points (95% CI –0.15% to –0.04%; $P=.001$), and medical expense per case significantly decreased by 5328.35 points (95% CI –10,387.10 to –269.60; $P=.04$).

XSL•FO
RenderX

Table 3. Comparative analysis of the individual and overall impact of categorization of hospital emergency capability (CHEC) policy effects on 4 critical time-sensitive diseases (N=9923).

Outcome	Change between before and after CHEC policy				Before CHEC policy	After CHEC policy	Multivariable model, β_1[a] (95% CI)	*P* value
	Acute ischemic stroke[b]	STEMI[c]	Septic shock[d]	Major trauma[e]				
Process quality								
Major diagnosis indicator, n (%)	–3.52	0.55	–2.28	–3.01	8682 (87.49)	8434 (84.99)	–0.21 (–0.29 to –0.13)	<.001
Diagnostic fees[f]	460.66	2746.6	–44.8	762.85	7166.94 (10,018.69)	7543.31 (10,833.26)	376.37 (92.42 to 660.33)	.01
Major treatment indicator, n (%)	0.62	2.07	–1.25	–3.12	4334 (43.68)	4272 (43.05)	–0.03 (–0.07 to 0.02)	.24
Medical orders per case	–0.93	–4.21	–9.67	–16.13	102.13 (109.19)	94.84 (101.07)	–7.29 (–10.09 to –4.48)	<.001
Upward transfer rate, n (%)	0.59	–1.24	–0.1	2.43	148 (1.49)	172 (1.73)	0.15 (–0.06 to 0.37)	.16
Outcomes								
30-d mortality, n (%)	–0.04	–1.52	–1.84	–1.27	1631 (16.44)	1508 (15.2)	–0.09 (–0.17 to –0.02)	.01
1-y mortality, n (%)	0.28	–0.83	–3.64	–0.58	3242 (32.67)	3041 (30.65)	–0.09 (–0.15 to –0.04)	.001
Total medical expense per case	3616.16	11,219.01	−11,059.09	−13,056.54	100,875.96 (192,912.94)	95,547.60 (176,487.79)	–5328.35 (–10,387.10 to –269.60)	.04

[a]Overall impact of the CHEC policy across all diseases.

[b]Acute ischemic stroke major diagnosis indicator: head image and major treatment indicator: intravenous tissue plasminogen activator thrombolysis.

[c]STEMI: ST-elevation myocardial infarction major diagnosis indicator: electrocardiography and major treatment indicator: percutaneous coronary intervention.

[d]Septic shock major diagnosis indicator: culture and major treatment indicator: antipathogen medication.

[e]Major trauma major diagnosis indicator: computed tomography, magnetic resonance imaging, or sonography study and major treatment indicator: rescue operation.

[f]Diagnostic fees: since Taiwan's National Health Insurance system operates on a global budget with reimbursement based on a point system, the actual monetary value of each point fluctuates. Currently, 1 National Health Insurance point is reimbursed at <NT $0.90 (US $0.0275) per point based on the latest exchange rate (1 NT $=US $0.03057 as of February 16, 2025).

Association of CHEC Policy With Processes and Outcome Quality in the 4 CTSDs

As presented in Table 4, model 1 examined the changes in indicators for individual diseases before and after the implementation of the CHEC policy, and the results show significant improvements in process quality measures. For AIS cases, following the implementation of the CHEC policy, there was a significant decrease in major diagnosis indicators by 0.23% points (95% CI –0.36% to –0.10%; *P*<.001). Conversely, the major treatment indicator experienced a significant increase of 0.57% points (95% CI 0.07%-1.07%; *P*=.03), and the upward transfer rate also significantly increased by 0.52% points (95% CI 0.02%-1.03%; *P*=.04). Moreover, there was a trend of increasing diagnostic fees, with a rise of 460.66 points (95% CI –3.44 to 924.76; *P*=.05). For STEMI cases, the diagnostic fees significantly increased by 2746.59 points (95% CI 1141.67-4351.51; *P*<.001). When examining septic shock, the major diagnosis indicator saw a significant decrease of 0.25% points (95% CI –0.37% to –0.12%; *P*<.001) following the introduction of the CHEC policy. Thirty-day mortality decreased by 0.11% (95% CI –0.20% to –0.02%; *P*=.02), and one-year mortality decreased by 0.15% (95% CI –0.22% to –0.07%; *P*<.001). In addition, medical orders significantly dropped by 9.67 items (95% CI –13.99 to –5.35; *P*<.001), and average medical expenses significantly decreased by 11,059.10 points (95% CI –18,603.60 to –3514.55; *P*=.004). Finally, regarding major trauma cases, after CHEC policy implementation, the average medical orders significantly decreased by 16.13 items (95% CI –25.32 to –6.94; *P*<.001).

XSL•FO
RenderX

Table 4. Association of categorization of hospital emergency capability (CHEC) policy with process and outcome quality in 4 critical time-sensitive diseases.

Critical time-sensitive disease	Change between before and after CHEC policy	Model 1[a]		Model 2[b]	
		β_2 (95% CI)	*P* value	β_3 (95% CI)	*P* value
Acute ischemic stroke[c] (n=2895)					
Major diagnosis indicator, n (%)	–3.52	–0.23 (–0.36 to –0.10)	<.001	–0.06 (–0.32 to 0.20)	.66
Diagnostic fees[d]	460.66	460.66 (–3.44 to 924.76)	.05	–302.19 (–1419.15 to 814.77)	.60
Major treatment indicator, n (%)	0.62	0.57 (0.07 to 1.07)	*.03*	0.77 (0.21 to 1.33)	.007
Medical orders per case	−0.93	–0.93 (–4.64 to 2.78)	.62	15.20 (5.28 to 25.11)	.003
Upward transfer rate, n (%)	0.59	0.52 (0.02 to 1.03)	.04	0.21 (–0.39 to 0.81)	.49
Short-term mortality (30 d), n (%)	–0.04	−0.01 (−0.25 to 0.23)	.95	0.11 (–0.27 to 0.49)	.57
Long-term mortality (365 d), n (%)	0.28	0.02 (−0.12 to 0.16)	.77	0.06 (–0.22 to.033)	.69
Total medical expense per case[h]	3616.16	3616.15 (–3524.26 to 10,756.56)	.32	16,672.69 (–3581.75 to 36,927.12)	.11
ST-segment elevation myocardial infarction[e](n=723)					
Major diagnosis indicator, n (%)	0.55	0.09 (–0.32 to 0.49)	.68	0.26 (–0.21 to 0.72)	.28
Diagnostic fees	2746.6	2746.59 (1141.67 to 4351.51)	<.001	1983.75 (84.28 to 3883.21)	*.04*
Major treatment indicator, n (%)	2.07	0.09 (–0.11 to 0.30)	.38	0.30 (–0.03 to 0.62)	.07
Medical orders per case	−4.21	–4.21 (–13.14 to 4.72)	.36	11.92 (–0.90 to 24.73)	.07
Upward transfer rate, n (%)	–1.24	–0.28 (–0.76 to 0.21)	.27	–0.59 (–1.18 to –0.001)	.049
Short-term mortality (30 d), n (%)	–1.52	–0.10 (–0.36 to 0.16)	.45	–0.02 (–0.37 to 0.41)	.92
Long-term mortality (365 d), n (%)	–0.83	−0.04 (−0.26 to 0.18)	.72	–0.01 (–0.33 to .032)	.97
Total medical expense per case	11,219.01	11,219.00 (–4953.90 to 27,391.90)	.17	24,275.54 (–640.71 to 49,191.78)	.06
Septic shock[f] (n=5441)					
Major diagnosis indicator, n (%)	–2.28	−0.25 (−0.37 to −0.12)	<.001	–0.08 (–0.33 to 0.18)	.56
Diagnostic fees	–44.8	–44.80 (–412.26 to 322.66)	.81	–807.65 (–1888.03 to 272.74)	.14
Major treatment indicator, n (%)	–1.25	−0.06 (−0.14 to 0.02)	.14	0.14 (–0.12 to 0.41)	.28
Medical orders per case	–9.67	–9.67 (–13.99 to –5.35)	<.001	6.45 (–3.70 to 16.61)	.21
Upward transfer rate, n (%)	–0.1	–0.27 (–0.93 to 0.38)	.41	–0.59 (–1.32 to 0.15)	.12
Short-term mortality (30 d), n (%)	–1.84	–0.11 (–0.20 to –0.02)	.02	–0.01 (–0.29 to 0.31)	.95
Long-term mortality (365 d), n (%)	–3.64	–0.15 (–0.22 to –0.07)	<.001	–0.11 (–0.36 to 0.13)	.36
Total medical expense per case	–11,059.09	–11,059.10 (–18,603.60 to –3514.55)	.004	1997.45 (–18,403.00 to 22,397.86)	.85
Major trauma[g] (n=864)					
Major diagnosis indicator, n (%)	–3.01	–0.17 (–0.39 to 0.05)	.14	N/A[h]	N/A
Diagnostic fees	762.85	762.85 (–253.13 to 1778.82)	.14	N/A	N/A
Major treatment indicator, n (%)	–3.12	–0.21 (–0.46 to 0.05)	.11	N/A	N/A
Medical orders per case	–16.13	–16.13 (–25.32 to –6.94)	<.001	N/A	N/A
Upward transfer rate, n (%)	2.43	0.31 (–0.02 to 0.65)	.06	N/A	N/A
Short-term mortality (30 d), n (%)	–1.27	–0.12 (–0.41 to 0.17)	.43	N/A	N/A
Long-term mortality (365 d), n (%)	–0.58	–0.03 (–0.27 to 0.20)	.77	N/A	N/A

XSL•FO
RenderX

Critical time-sensitive disease	Change between before and after CHEC policy	Model 1[a]		Model 2[b]	
		β_2 (95% CI)	*P* value	β_3 (95% CI)	*P* value
Total medical expense per case	−13,056.54	–13,056.50 (–32,010.60 to 5897.53)	.18	N/A	N/A

[a]Model 1 compares the specific disease differences between before and after CHEC policy implementation.

[b]Model 2: model-adjusted estimates for an interaction between a binary measure of CHEC policy (ie, postimplementation vs preimplementation) and critical time-sensitive diseases compared with major trauma (eg, acute ischemic stroke vs major trauma, ST-segment elevation myocardial infarction vs major trauma, and septic shock vs major trauma).

[c]Acute ischemic stroke major diagnosis indicator: head image and major treatment indicator: intravenous tissue plasminogen activator thrombolysis.

[d]Diagnostic fees: since Taiwan's National Health Insurance system operates on a global budget with reimbursement based on a point system, the actual monetary value of each point fluctuates. Currently, 1 National Health Insurance point is reimbursed at <NT $0.90 (US $0.0275) per point based on the latest exchange rate (NT $1=US $0.03057 as of February 16, 2025).

[e]STEMI: ST-segment elevation myocardial infarction major diagnosis indicator: electrocardiography and major treatment indicator: percutaneous coronary intervention.

[f]Septic shock major diagnosis indicator: culture and major treatment indicator: antipathogen medication.

[g]Major trauma major diagnosis indicator: computed tomography, magnetic resonance imaging, or sonography study and major treatment indicator: rescue operation.

[h]N/A: data not applicable as major trauma cases were the reference group.

In model 2, results from the GEE model highlighted the CHEC policy's varied effects across different diseases. A positive group-by-disease interaction β_3 coefficient indicated that the outcome changes for that disease was greater than the reference group (Table 4). Compared to major trauma cases, AIS cases exhibited a significant increase in the major treatment indicator (interaction β coefficient=0.77; 95% CI 0.21-1.33; *P*=.007) and medical orders (interaction β coefficient=15.20; 95% CI 5.28-25.11; *P*=.003) between before and after CHEC policy implementation. Meanwhile, STEMI cases demonstrated a significant increase in diagnostic fees (interaction β coefficient=1983.75; 95% CI 84.28-3883.21; *P*=.04) and a significant decrease in upward transfer rate (interaction β coefficient=−0.59; 95% CI –1.18 to –0.001; *P*=.049) than the major trauma cases. Moreover, there were trends toward increasing major treatment indicators (interaction β coefficient=0.30; 95% CI –0.03 to 0.62; *P*=.07), medical orders (interaction β coefficient=11.92; 95% CI –0.90 to 24.73; *P*=.07), and medical expense (interaction β coefficient=24,275.54; 95% CI –640.71 to 4,991,991.78; *P*=.06), although these were not statistically significant. Compared to major trauma cases, no significant change was observed in either process or outcome quality indicators for septic shock cases.

Discussion

Principal Findings

The CHEC policy was universally implemented across Taiwan's emergency medical service systems. Evaluating the impact of such a policy on a nationwide population presents significant challenges, primarily due to the absence of a control group. This limitation restricts the analysis to pre- and postimplementation comparisons, complicating the understanding of how the policy may influence shifts in various diseases. Despite these challenges, our study aimed to analyze the policy's differential effects on time-sensitive conditions thoroughly. By examining disease-specific guidelines, we identified the groups most affected by the policy and those serving as relatively unaffected counterfactuals, offering valuable insights into the policy's impact. To address these challenges, we used a DID design combined with PSM, using major trauma cases as a reference group. This approach ensured that baseline covariates were balanced between before and after policy periods, allowing us to investigate the policy's effects while controlling for preexisting differences in participant and hospital characteristics. Our analysis revealed 2 major findings. First, the CHEC policy effectively improved system efficiency and patient outcomes, resulting in significant reductions in medical orders (ie, –7.29 items per case), short- and long-term mortality rates (ie, –0.09% each), and total medical expenses (ie, –5328.35 points per case), despite a modest increase in diagnostic fees (ie, 376.37 points). Second, we observed unintended "policy spotlight effects," where conditions with time-based surveillance indicators, such as AIS and STEMI, showed disproportionately greater improvements than conditions without such indicators. Specifically, AIS cases experienced significant increases in major treatments (β=0.77) and medical orders (β=15.20), while STEMI cases demonstrated increased diagnostic fees (β=1983.75) and decreased upward transfer rates (β=–0.59) relative to major trauma cases. These findings highlight the varied impacts of the policy based on the presence or absence of time-based monitoring indicators. Thus, the effectiveness and efficiency of the CHEC policy underscore its dual ability to lower costs while improving patient outcomes. However, why has the CHEC policy significantly reduced medical costs and mortality rates, even though the major diagnosis indicator has declined, and no substantial changes have been observed in treatment indicators? The subsequent sections will delve further into diseases' individual and interactive effects to provide analysis.

Distinction Between the Hawthorne and Policy Spotlight Effects

To distinguish health care providers' behaviors influenced by the Hawthorne effect or policy spotlight effects, we selected AIS and STEMI, which have well-established guidelines and

XSL•FO
RenderX

time-based quality indicators under the CHEC policy. In contrast, septic shock, a disease with well-established guidelines but no specific time-based quality indicators, and major trauma, lacking both well-established guidelines and time-based quality indicators, were chosen as reference groups (Table S2 in Multimedia Appendix 1).

After implementing the CHEC policy, we hypothesized that the observed responses might indicate varying levels of awareness among emergency care providers [24]. The Hawthorne effect highlights that individuals' productivity in experimental settings may increase due to the simple fact that they are being observed. This phenomenon underscores the influence of human attention and intervention on behavior [25]. The policy spotlight effects may be intensified by factors such as time constraints, ambiguous symptom patterns, and time-based quality surveillance indicators, prompting emergency care providers to unconsciously adopt selective behaviors concentrating on specific diseases according to policy targets [25]. Consequently, diseases not prioritized by the policy, such as septic shock and major trauma in this study, may experience a significant decline in process quality and a reduction in medical orders on an individual disease basis (Table 3). When using major trauma as the reference group, both AIS and STEMI group showed increases in major treatments, medical orders, diagnostic fees, and medical expenses. In contrast, no significant changes were observed in the septic shock group (Table 4).

When examining the effects of the policy on individual diseases, we observed distinct patterns. A decrease in diagnostic indicators was noted for AIS cases, likely due to increased upward transfer rates, as hospital physicians appeared more inclined to transfer patients with AIS to higher-level hospitals to enhance access to thrombolytic treatment [26]. This aligns with a significant increase in primary treatment indicators. For STEMI cases, there was an increase in diagnostic fees and a notable rise in diagnostic and treatment indicators, coupled with decreased transfer rates, suggesting a more intensive treatment approach that correlates with reduced mortality rates. These findings highlight the differential impacts of the policy, with AIS and STEMI cases benefiting from improvements in process and treatment quality due to the presence of time-based indicators and targeted monitoring. In contrast, septic shock and major trauma cases, which lack such indicators, showed either no significant changes or potential deterioration in quality metrics. These contrasting trends between diseases with and without time-based monitoring suggest the influence of policy spotlight effects. Under the CHEC policy, AIS and STEMI cases have demonstrated marked improvements, while septic shock and major trauma cases remain relatively unaffected, underscoring the importance of targeted policy monitoring in driving quality improvements.

While the Hawthorne effect typically manifests as generalized behavioral improvements across most observed conditions due to the awareness of being monitored, the implementation of the CHEC policy was explicitly designed to optimize emergency medical capacity through enhanced adherence to evidence-based guidelines, with a particular emphasis on reducing unnecessary medical orders and improving emergency care efficiency. The observed outcomes demonstrated significant improvements, including reductions in medical orders and per-case expenses, concurrent with decreased short- and long-term mortality rates, which are the overall impact of CHEC policy effects on the 4 CTSDs. These observed improvements may be attributed to the Hawthorne effect, wherein the awareness of being observed leads to enhanced performance among health care providers.

Both AIS and STEMI cases appeared to be influenced by the policy spotlight effect, and they showed opposite results in diagnostic indicators and upward transfer rates. These could potentially be attributed to the inherent differences between AIS and STEMI cases. For instance, only about 1% to 2% of all patients with AIS underwent primary treatment with intravenous thrombolytic agents, which is far lower than the STEMI major treatment rate. Unexpectedly, in less policy spotlight affected groups, such as septic shock and major trauma, despite significant decreases in primary diagnostic indicators, medical orders, and diagnostic fees, there were unexpected reductions in 30-day and 1-year mortality rates and medical expenses per event. These unintended consequences might resonate with the "less is more" initiative [27], suggesting that curbing excessive diagnoses and avoiding unnecessary procedures could improve patient outcomes and decrease costs [28,29].

Policy Implications

Many emergency care policies implement time-based criteria [30-32], such as Australia's 4-hour rule [30] and the United Kingdom's 4-hour standard [32]. Australia's experience showed that an emergency care policy using time-based criteria could improve emergency congestion without increasing the rate of ED revisits. However, in New Zealand, a policy in effect from 2006 to 2012 dictated that emergency patients must be hospitalized, transferred, or discharged within 6 hours of visiting the ED. After emergency care policy intervention, the length of ED stay decreased while the treatment outcomes of acute myocardial infarction, severe septic shock, and acute appendicitis did not improve significantly [31]. Similarly, after the Canadian Emergency Observation Reduction Program implementation, the length of ED stay decreased, while the treatment quality indicators for acute myocardial infarction, asthma, and upper limb fractures could only be treated in time for the abovementioned time-sensitive diseases during the noncongested emergency period [33].

Emergency care quality is closely related to the practices of medical care providers [12]. Policy makers must reconsider the conventional emphasis on time-based process indicators. This approach may have unintended consequences for diseases that are outside the "spotlight" of rigid time-sensitive evaluation. Instead, a broader and more nuanced evaluation of emergency care quality is needed to incorporate the complexity and ambiguity of various time-sensitive diseases that emergency care providers often manage [12]. We propose replacing time-based indicators with a broader set of multidimensional performance-based indicators, such as those exemplified in the National Health Service Best Practice Tariff policy [34]. This shift toward "best practice" can create a more flexible approach that goes beyond merely relying on time-based or diagnosis-based practices [34]. Such a transition is crucial, as an excessive reliance on diagnostic tests may lead to ED

XSL•FO
RenderX

crowding, a decline in emergency care quality, and increased safety issues [28].

Strengths

This study used a robust DID method within a GEE framework to evaluate the impact of the CHEC policy. The analysis compared pre- and postintervention periods and distinguished treatment and control groups using a categorical disease variable. Interaction terms captured differential policy effects across disease groups, while GEE accounted for correlated data from repeated measures, ensuring robust variance estimation. The estimated coefficients quantified both overall and disease-specific policy impacts, providing a reliable and nuanced evaluation of the CHEC policy. Our study also examined the behaviors of emergency care providers under time constraints and their interactions with strict time-based quality surveillance indicators and adherence to "get with the guidelines" protocols, revealing policy spotlight effects.

Limitations

This study has several limitations. First, the data were derived from a secondary dataset of insurance claims rather than a randomized controlled trial, which may limit causal inferences. Second, the analysis lacked detailed information on time-related quality indicators, such as door-to-evaluation and door-to-treatment times, and critical emergency care metrics. Third, the effectiveness of emergency medical care often depends on collaboration between emergency and consulting physicians; however, this study did not explicitly evaluate the dynamics of their interaction. Future research should examine the coordination and teamwork between these physician groups to understand their impact on care quality better.

Conclusions

Our study reveals that CHEC policy implementation demonstrates a dual capability to reduce costs and improve patient outcomes. The policy spotlight effects result in a disproportional improvement in disease guideline adherence and process quality of CTSDs with time-based surveillance indicators. In contrast, disease entities not fully encompassed in the surveillance indicators may be jeopardized by decreasing diagnosis and treatment process quality.

Acknowledgments

The authors would like to acknowledge the Taiwan National Health Insurance Research Database, provided by the National Health Insurance Administration and managed by the National Health Research Institutes. They would also like to thank the Center for Public Health, Department of Education and Research, Taipei City Hospital, for providing administrative and research grant support. The authors extend their deepest gratitude to the esteemed members of the dissertation committee for their invaluable guidance and support throughout their research journey. The concept of this paper, titled "The Structure, Process, and Outcome Quality Impacts of the Categorization of Hospital Emergency Capability Policy," was significantly shaped by their insightful feedback and thoughtful encouragement, for which they are profoundly grateful.

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Generative AI tools (eg, ChatGPT by OpenAI) were used to assist with language refinement and reformatting text for clarity. All AI-generated content was thoroughly reviewed, validated, and edited by the authors to ensure accuracy, scientific rigor, and adherence to academic standards. No content was directly incorporated without author verification, ensuring the integrity of the manuscript remains uncompromised.

Data Availability

The data that support the findings of this study are available from the Taiwan National Health Insurance Research Database, but restrictions apply to the availability of these data, which were used under license for this study, and so are not publicly available. Data are, however, available at the academic request and with the permission of the Taiwan National Health Insurance Administration.

Authors' Contributions

CYL took the lead in conceptualizing the study, drafting the original manuscript, and conducting the formal data analysis. In addition, CYL verified the underlying data presented in the paper. CCL contributed to the study design, data curation, and formal data analysis and oversaw the data collection process. Both CCL and YTH ensured the accuracy of the data analysis and interpretation, as well as verified the paper's underlying data. YCL supervised the study, validated the findings, and played a significant role in reviewing and editing the manuscript. All authors contributed to the development of the study concept and design, as well as the analysis and interpretation of data and the preparation of the manuscript. All authors have approved the final version of the manuscript and take collective responsibility for all aspects of the work, pledging to investigate and resolve any concerns related to the accuracy or integrity of the work.

Conflicts of Interest

None declared.

XSL•FO
RenderX

Multimedia Appendix 1
Propensity score–matched patient characteristics for 4 time-sensitive diseases before and after categorization of hospital emergency capability (CHEC) policy implementation, association of CHEC policy with process and outcome quality in 4 critical time-sensitive diseases, and the distinction between the Hawthorne and policy spotlight effects.
[DOC File , 157 KB - ijmr_v14i1e54651_app1.doc]

Multimedia Appendix 2
Association of CHEC policy with process and outcomes quality in four critical time-sensitive diseases.
[DOCX File , 25 KB - ijmr_v14i1e54651_app2.docx]

References

1. AHRQ quality indicators—guide to prevention quality indicators: hospital admission for ambulatory care sensitive conditions. Agency for Healthcare Research and Quality. 2001. URL: https://www.ahrq.gov/downloads/pub/ahrqqi/pqiguide.pdf [accessed 2025-02-13]
2. Combes JR, Arespacochaga E. Appropriate use of medical resources. American Hospital Association's Physician Leadership Forum. 2013. URL: https://www.aha.org/system/files/content/13/appropusewhiteppr.pdf [accessed 2025-02-13]
3. Panchavati S, Lam C, Zelin NS, Pellegrini E, Barnes G, Hoffman J, et al. Retrospective validation of a machine learning clinical decision support tool for myocardial infarction risk stratification. Healthc Technol Lett 2021 Dec 31;8(6):139-147 [FREE Full text] [doi: 10.1049/htl2.12017] [Medline: 34938570]
4. Daley S, Braimah J, Sailor S, Kongable GL, Barch C, Rapp K, et al. Education to improve stroke awareness and emergent response. The NINDS rt-PA Stroke Study Group. J Neurosci Nurs 1997 Dec;29(6):393-396. [doi: 10.1097/01376517-199712000-00009] [Medline: 9479661]
5. Fesmire FM, Brady WJ, Hahn S, Decker WW, Diercks DB, Ghaemmaghami CA, et al. Clinical policy: indications for reperfusion therapy in emergency department patients with suspected acute myocardial infarction. American College of Emergency Physicians Clinical Policies Subcommittee (Writing Committee) on reperfusion therapy in emergency department patients with suspected acute myocardial infarction. Ann Emerg Med 2006 Oct;48(4):358-383. [doi: 10.1016/j.annemergmed.2006.07.006] [Medline: 16997672]
6. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. N Engl J Med 2001 Nov 08;345(19):1368-1377. [doi: 10.1056/NEJMoa010307] [Medline: 11794169]
7. Carr BG, Kilaru AS, Karp DN, Delgado MK, Wiebe DJ. Quality through coopetition: an empiric approach to measure population outcomes for emergency care-sensitive conditions. Ann Emerg Med 2018 Sep;72(3):237-245 [FREE Full text] [doi: 10.1016/j.annemergmed.2018.03.004] [Medline: 29685369]
8. Mehrotra A, Sklar DP, Tayal VS, Kocher KE, Handel DA, Myles Riner R. Important historical efforts at emergency department categorization in the United States and implications for regionalization. Acad Emerg Med 2010 Dec;17(12):e154-e160 [FREE Full text] [doi: 10.1111/j.1553-2712.2010.00931.x] [Medline: 21122008]
9. Haupt MT, Bekes CE, Brilli RJ, Carl LC, Gray AW, Jastremski MS, et al. Guidelines on critical care services and personnel: recommendations based on a system of categorization of three levels of care. Crit Care Med 2003 Nov;31(11):2677-2683. [doi: 10.1097/01.CCM.0000094227.89800.93] [Medline: 14605541]
10. Glickman SW, Lytle BL, Ou FS, Mears G, O'Brien S, Cairns CB, et al. Care processes associated with quicker door-in-door-out times for patients with ST-elevation-myocardial infarction requiring transfer: results from a statewide regionalization program. Circ Cardiovasc Qual Outcomes 2011 Jul;4(4):382-388. [doi: 10.1161/CIRCOUTCOMES.110.959643] [Medline: 21712523]
11. Govindarajan P, Shiboski S, Grimes B, Cook LJ, Ghilarducci D, Meng T, et al. Effect of acute stroke care regionalization on intravenous alteplase use in two urban counties. Prehosp Emerg Care 2020 Nov 11;24(4):505-514. [doi: 10.1080/10903127.2019.1679303] [Medline: 31599705]
12. Rubenstein LV, Mittman BS, Yano EM, Mulrow CD. From understanding health care provider behavior to improving health care: the QUERI framework for quality improvement. Med Care 2000;38(6):I-129-I-141. [doi: 10.1097/00005650-200006001-00013]
13. Gilovich T, Savitsky K. The spotlight effect and the illusion of transparency: egocentric assessments of how we are seen by others. Curr Dir Psychol Sci 1999 Dec 01;8(6):165-168. [doi: 10.1111/1467-8721.00039]
14. Salisbury KM. National and state policies influencing the care of children affected by AIDS. Child Adolesc Psychiatr Clin N Am 2000 Apr;9(2):425-449. [doi: 10.1016/s1056-4993(18)30128-7]
15. Savitsky K, Epley N, Gilovich T. Do others judge us as harshly as we think? Overestimating the impact of our failures, shortcomings, and mishaps. J Pers Soc Psychol 2001 Jul;81(1):44-56. [doi: 10.1037//0022-3514.81.1.44] [Medline: 11474724]
16. Lin LY, Warren-Gash C, Smeeth L, Chen PC. Data resource profile: the National Health Insurance Research Database (NHIRD). Epidemiol Health 2018;40:e2018062 [FREE Full text] [doi: 10.4178/epih.e2018062] [Medline: 30727703]

XSL•FO
RenderX

17. Lin CY, Lee YC. Effectiveness of hospital emergency department regionalization and categorization policy on appropriate patient emergency care use: a nationwide observational study in Taiwan. BMC Health Serv Res 2021 Jan 06;21(1):21 [FREE Full text] [doi: 10.1186/s12913-020-06006-7] [Medline: 33407444]
18. Hospital Emergency Capability Categorization. Ministry of Health and Welfare Taiwan. URL: https://dep.mohw.gov.tw/DOMA/np-980-106.html [accessed 2025-02-27]
19. Shafi S, Nathens AB, Cryer HG, Hemmila MR, Pasquale MD, Clark DE, et al. The trauma quality improvement program of the American College of Surgeons Committee on Trauma. J Am Coll Surg 2009 Oct;209(4):521-30.e1. [doi: 10.1016/j.jamcollsurg.2009.07.001] [Medline: 19801325]
20. Cox S, Smith K, Currell A, Harriss L, Barger B, Cameron P. Differentiation of confirmed major trauma patients and potential major trauma patients using pre-hospital trauma triage criteria. Injury 2011 Sep;42(9):889-895. [doi: 10.1016/j.injury.2010.03.035] [Medline: 20430387]
21. Dobaria V, Kwon OJ, Hadaya J, Sanaiha Y, Sareh S, Aguayo E, et al. Impact of center volume on outcomes of surgical repair for type A acute aortic dissections. Surgery 2020 Jul;168(1):185-192. [doi: 10.1016/j.surg.2020.04.007] [Medline: 32507629]
22. Asplin BR, Magid DJ, Rhodes KV, Solberg LI, Lurie N, Camargo CAJ. A conceptual model of emergency department crowding. Ann Emerg Med 2003 Aug;42(2):173-180. [doi: 10.1067/mem.2003.302] [Medline: 12883504]
23. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res 2011 May 31;46(3):399-424 [FREE Full text] [doi: 10.1080/00273171.2011.568786] [Medline: 21818162]
24. Lied TR, Kazandjian VA. A Hawthorne strategy: implications for performance measurement and improvement. Clin Perform Qual Health Care 1998;6(4):201-204. [Medline: 10351289]
25. Franke RH, Kaul JD. The Hawthorne experiments: first statistical interpretation. Am Sociol Rev 1978 Oct;43(5):623-643. [doi: 10.2307/2094540]
26. Zachrison KS, Amati V, Schwamm LH, Yan Z, Nielsen V, Christie A, et al. Influence of hospital characteristics on hospital transfer destinations for patients with stroke. Circ Cardiovasc Qual Outcomes 2022 May;15(5):e008269. [doi: 10.1161/CIRCOUTCOMES.121.008269] [Medline: 35369714]
27. Wegwarth O, Gigerenzer G. Less is more: overdiagnosis and overtreatment: evaluation of what physicians tell their patients about screening harms. JAMA Intern Med 2013 Dec 09;173(22):2086-2087. [doi: 10.1001/jamainternmed.2013.10363] [Medline: 24145597]
28. Kawano T, Nishiyama K, Hayashi H. Execution of diagnostic testing has a stronger effect on emergency department crowding than other common factors: a cross-sectional study. PLoS One 2014 Oct 13;9(10):e108447 [FREE Full text] [doi: 10.1371/journal.pone.0108447] [Medline: 25310089]
29. Hoot NR, Aronsky D. Systematic review of emergency department crowding: causes, effects, and solutions. Ann Emerg Med 2008 Aug;52(2):126-136 [FREE Full text] [doi: 10.1016/j.annemergmed.2008.03.014] [Medline: 18433933]
30. Ngo H, Forero R, Mountain D, Fatovich D, Man WN, Sprivulis P, 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 Mar 14;13(3):e0193902 [FREE Full text] [doi: 10.1371/journal.pone.0193902] [Medline: 29538401]
31. Jones P, Le Fevre J, Harper A, Wells S, Stewart J, Curtis E, et al. Effect of the shorter stays in emergency departments time target policy on key indicators of quality of care. N Z Med J 2017 May 12;130(1455):35-44. [Medline: 28494476]
32. Cooke MW. Reforming the UK emergency care system. Emerg Med J 2003 Mar;20(2):113-114 [FREE Full text] [doi: 10.1136/emj.20.2.113] [Medline: 12642515]
33. Vermeulen MJ, Guttmann A, Stukel TA, Kachra A, Sivilotti ML, Rowe BH, et al. Are reductions in emergency department length of stay associated with improvements in quality of care? A difference-in-differences analysis. BMJ Qual Saf 2016 Jul 13;25(7):489-498 [FREE Full text] [doi: 10.1136/bmjqs-2015-004189] [Medline: 26271919]
34. Copas DP, Moran CG. Major trauma care in England: changing the state of a nation's healthcare system. Bone Joint 360 2014 Apr;3(2):2-5. [doi: 10.1302/2048-0105.32.360217]

Abbreviations

AIS: acute ischemic stroke
CHEC: categorization of hospital emergency capability
CTSD: critical time-sensitive disease
DID: difference-in-differences
ED: emergency department
GEE: generalized estimating equation
ICD-9-CM: International Classification of Diseases, Ninth Revision, Clinical Modification
LHID2005: Longitudinal Health Insurance Database
PSM: propensity score matching
STEMI: ST-segment elevation myocardial infarction

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 26.11.23; peer-reviewed by A Debbarma, T Cinar, M Al Zoubi; comments to author 06.01.25; revised version received 20.01.25; accepted 23.01.25; published 25.03.25.

Please cite as:
Lin CY, Liu CC, Huang YT, Lee YC
Policy Spotlight Effects on Critical Time-Sensitive Diseases: Nationwide Retrospective Cohort Study on Taiwan's Hospital Emergency Capability Categorization Policy
Interact J Med Res 2025;14:e54651
URL: https://www.i-jmr.org/2025/1/e54651
doi:10.2196/54651
PMID:

©Chih-Yuan Lin, Chih-Ching Liu, Yu-Tung Huang, Yue-Chune Lee. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 25.03.2025. 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

Streamlining Ophthalmic Documentation With Anonymized, Fine-Tuned Language Models: Feasibility Study

Sebastian Arens[1], BSc, MD; Quang Vinh Ngo[1], Dr Med; Anna Richling[1], MD; Lucas Stürzbecher[1], MD; Daniel Böhringer[1], Prof Dr Med; Thomas Reinhard[1], Prof Dr Med; Felix Heilmeyer[2], MD

[1]Eye Center, University Medical Center Freiburg, Freiburg im Breisgau, Baden Wuerttemberg, Germany

[2]Center for Digitization in Medicine, University Medical Center Freiburg, Freiburg im Breisgau, Baden Wuerttemberg, Germany

Corresponding Author:
Sebastian Arens, BSc, MD
Eye Center
University Medical Center Freiburg
Killianstrasse 5
Freiburg im Breisgau, Baden Wuerttemberg, 79106
Germany
Phone: 49 76127040010
Email: sebastian.arens@uniklinik-freiburg.de

Abstract

Background: The growing administrative burden on clinicians, particularly in medical documentation, contributes to burnout and may compromise patient safety. Recent advancements in generative artificial intelligence (AI) offer a promising solution to improve documentation processes and address these challenges.

Objective: This study aims to evaluate the feasibility of using a fine-tuned OpenAI Curie model to automate the generation of medical report summaries (epicrises) in ophthalmology. By assessing the model's performance through human and automated evaluations, this study seeks to determine its potential for reducing clinician workload while ensuring accuracy, usefulness, and compliance with regulatory requirements.

Methods: A data set of around 60,000 anonymized medical letters was created using a custom algorithm to comply with General Data Protection Regulation guidelines. The Curie model was fine-tuned on this data set to generate epicrises from medical histories, diagnoses, and findings. The performance evaluation involved various human assessments and automated evaluations from 2 large language models (LLMs).

Results: In the clinical context, 49.9% (384/769) of epicrises were evaluated as helpful or excellent, whereas only 25% (194/769) were considered disturbing. In a human (manual) evaluation, formal correctness was rated significantly higher than the neutral midpoint of 2.5 on the 4-point rating scale, as determined by a 1-sample Wilcoxon signed-rank test (mean 3.59, SD 0.85; W=1686; P<.001). Using paired t tests, we found a significant reduction in time, as correcting an AI epicrisis was faster than manually writing one (mean 109.52, SD 53.30 vs mean 54.25, SD 63.34 s; t_{68}=3.39; P<.01). While medical accuracy and usefulness showed positive trends, these did not reach statistical significance when compared to the neutral midpoint (for medical accuracy, W=7456; P=.08), for usefulness, W=7652.5; P=.18). Epicrises generated or corrected with AI were significantly shorter than manually written ones (mean 330.43, SD 115.42 vs mean 501.07, SD 243.50 characters; t_{68}=–6.10; P<.001). Automated LLM assessments showed alignment with human ratings, with over 52% (356/679) and 66% (489/743) of responses in the top agreement categories, respectively. This supports overall consistency, though the comparison remains a proof of concept given methodological limitations.

Conclusions: Our study demonstrates the technical and practical feasibility of introducing fine-tuned commercial LLMs into clinical practice. The AI-generated epicrises were formally and clinically correct in many cases and showed time-saving potential. While medical accuracy and usefulness varied across cases and should be focused on in further developments, a significant workload reduction is likely. Our anonymization process showed that regulatory challenges in the context of AI with patient data can effectively be dealt with. In summary, this study highlights the promise of transformer-based LLMs in reducing administrative tasks in health care. It outlines a pipeline for integrating LLMs into European Union clinical practice, emphasizing the need for careful implementation to ensure efficiency and patient safety.

(Interact J Med Res 2025;14:e72894) doi:10.2196/72894

XSL•FO
RenderX

KEYWORDS

artificial intelligence; large language models; medical reports; epicrisis generation; documentation automation

Introduction

In modern health care, administrative tasks have seen a sharp rise, escalating clinician workloads and increasing the risk of burnout and potential patient hazards due to an overload of work. [1,2] This trend has led to a critical demand within the medical community for effective and reliable assistance with documentation tasks. In parallel, the world of artificial intelligence (AI) has witnessed the development of large language models (LLMs) by organizations such as OpenAI and Anthropic and big tech companies like Google. These models have demonstrated their capability for generating coherent, human-like text that closely resembles professionally written content. However, out-of-the-box solutions like ChatGPT and similar models are, in many cases, not viable for immediate implementation in medical settings. This is due to several factors, including "refusal" due to built-in safety checks, the lack of alignment with local treatment guidelines, and the absence of detailed domain-specific knowledge [3,4].

To address these limitations, the concept of fine-tuning LLMs has emerged. Fine-tuning involves training existing language models on specialized data sets to enhance their performance in specific domains. Haghigi et al [5] demonstrated that training LLMs with specialized ophthalmic data improves their effectiveness in clinical decision-making and patient education. Their EYE-Llama models, pretrained on ophthalmology-specific data sets and fine-tuned on diverse question-answering data sets, outperformed baseline Llama 2 and ChatDoctor models, achieving results comparable to ChatGPT.

Van Veen et al [6] demonstrated that LLMs can outperform medical experts in summarizing clinical text across tasks like radiology reports, patient questions, and progress notes. Their study found that in most cases, summaries generated by adapted LLMs were equivalent or superior to those produced by human experts. This suggests that integrating LLMs into clinical workflows could significantly reduce the documentation burden on clinicians, thus allowing them to dedicate more time to patient care.

This study examines the feasibility of using adapted LLMs to automate medical report writing, specifically focusing on generating epicrises, or summaries. In our ophthalmology reports, there are three main sections: (1) patient history and diagnosis, (2) clinical findings, and (3) the epicrisis, which synthesizes the key information. We identified the epicrisis as the most suitable target for LLM-based assistance, as it synthesizes and consolidates information from the earlier sections. This reduces the risk of hallucinations by the model since the content is already documented and can be cross-referenced. In contrast, sections such as history and findings present more uncertainty and are more prone to variability or missing data, making them vulnerable to confabulation [7].

Epicrises are the most promising targets for automating the generation of medical reports. Heilmeyer et al [8] already provided valuable insights into fine-tuning LLM on premises for epicrisis generation to fulfill a similar task. To achieve this, different LLMs were fine-tuned on local servers to generate parts of medical reports, with varying degrees of success. We hypothesized that large commercial LLMs, such as OpenAI, hosted on capable infrastructure could outperform the local LLMs with limited resources and may be more accessible to sites with lower local resources. To our knowledge, OpenAI does not fully disclose the infrastructure on which the models are hosted, but it is known to use the Microsoft Azure Cloud and have access to a state-of-the-art graphics processing unit (GPU), such as NVIDIA H100, and server infrastructure [9].

Transferring medical reports to the United States for the purpose of improving medical report writing using an LLM poses significant challenges due to stringent legal and regulatory requirements. The General Data Protection Regulation (GDPR) in the European Union (EU) sets high standards for the protection of personal data, including medical information. According to the GDPR, the transfer of personal data to a third country, such as the United States, is only legal under certain conditions. Two primary approaches can be taken to make this transfer feasible: anonymizing the data beforehand or using Standard Contractual Clauses (SCCs) in combination with obtaining individual consent [8,10-12].

First, fully anonymizing the text under the legal basis of "legitimate interest" and validating the results as part of a Data Protection Impact Assessment (DPIA) is one approach. Once the data are anonymized, they no longer fall under the stringent regulations governing personal data, thus facilitating the transfer.

Alternatively, the use of SCCs or Binding Corporate Rules (BCRs), coupled with obtaining individual consent, can provide a lawful basis for the transfer. The invalidation of the Privacy Shield framework by the European Court of Justice (Schrems II decision) necessitates these legal mechanisms, along with additional safeguards, to ensure an adequate level of data protection [12]. Furthermore, ensuring compliance with the Health Insurance Portability and Accountability Act (HIPAA) is essential for any US entity handling health information. These regulations require comprehensive risk assessments, stringent data protection measures, and maintaining transparency with data subjects [13,14].

By either fully anonymizing the data before transfer or utilizing SCCs in combination with individual consent, the process of leveraging LLMs for enhancing medical report writing can be successfully navigated within the regulatory framework.

Other groups have already investigated approaches to anonymize clinical report data. Wiest et al [15] presented an LLM-based anonymization pipeline that can effectively deidentify medical free-text data, facilitating secure data sharing for medical research. This approach used a different locally installed LLM to anonymize clinical text data.

Therefore, the primary aim of this feasibility study was to evaluate the potential of a fine-tuned LLM, specifically

XSL•FO
RenderX

OpenAI's Curie model, to automate the generation of ophthalmic epicrises. We hypothesized that such a model, when fine-tuned on a large, anonymized data set of clinical reports, could generate accurate and clinically useful summaries, thereby demonstrating the potential to reduce clinician documentation workload. This study specifically sought to assess the feasibility of this approach, the quality of the generated epicrises through both human and automated evaluations, and potential time savings, all while ensuring compliance with data protection regulations like GDPR.

Methods

Overview

Our approach involved obtaining quality training data from a large pool of consecutive medical letters available in our electronic database, which numbered close to 60,000. The data had no inclusion criteria to avoid introducing bias and to reflect real-world medical letters. Only high-quality medical letters that were reviewed by both a resident and an attending physician were included. Incomplete letters with missing fields were excluded.

Anonymization Process

The following custom anonymization algorithm was developed and applied to all letters. In the first step, we extracted the data using an SQL query that selectively recalled the content of the letters but not any identifying data, such as names and addresses. The patient ID, however, was kept alongside the textual data for the second step. Here, we handled the rare cases where identifying information had been placed within the letters' contents, such as a reference to the patient's name and/or precise temporal information. This was done by explicitly replacing the patient's first name and surname with specific placeholders, using the patient ID to look up all identifying words to be eliminated from the document. The third step involved applying several regular expressions to capture date/time values, third-party names, and postal addresses. All occurrences were substituted with specific constant placeholders to preserve the semantic context and grammar.

This approach ensured that there was only a comparably small remaining text corpus containing only the information needed for the individual visit and not the whole patient history. This makes the anonymization process more consistent. Deanonymization in this case can only be assumed to be possible for highly individual cases and rare diseases. To ensure the robustness of our anonymization process during a DPIA, we conducted a rigorous evaluation. A randomly selected sample of 300 anonymized letters was independently reviewed by 2 expert investigators who were blinded to the original content. Neither reviewer was able to identify any residual personal data, demonstrating the algorithm's effectiveness in protecting patient privacy. Following this successful validation, our data protection officer approved the transfer of the anonymized data set to OpenAI.

We opted for this custom-made approach because it offers excellent reproducibility and explainability, giving us a high level of control over the anonymization process to ensure its quality. While other anonymization methods, including those using LLMs, exist, they do not provide the same degree of control, making our approach more reliable for our needs [15].

Training and Fine-Tuning Process

OpenAI offers self-service fine-tuning for selected models from the GPT-3 family. We opted for the second powerful base model named "Curie" at the time of training since we wanted to assess the feasibility of our approach and not get a false-negative outcome due to selecting a base model with insufficient capabilities [16]. The training data set comprised a total of 90 million tokens, which were derived from approximately 60,000 medical letters.

We conducted a simple completion training in which the full medical letter, excluding the epicrisis, served as the prompt and the epicrisis as the corresponding completion. We did not use any instructions during training.

The fine-tuning of our model was achieved by using the OpenAI fine-tuning API via the official command line tools. The fine-tuning methodology of OpenAI is not publicly disclosed to the best of our knowledge. However, the low pricing and 24/7 availability of the fine-tuned model make it likely that some kind of adaptor technology is used instead of fully fine-tuning the base model. Hyperparameters, such as learning rate and batch size, are chosen automatically by the OpenAI command line tools.

The overall cost of fine-tuning totaled approximately 500 € (US $581). The fine-tuned model was hosted by OpenAI and accessed via a private key using the official API during the study period. At the time of publishing this article, the model has now been deactivated because its base model was depreciated by OpenAI, although the API remains operational [16].

Integration Into a Testing Environment for a Global Assessment

We built a custom evaluation software tool. In this step, the physicians were prompted to evaluate the quality of sequential model-generated epicrises for anonymized data from real patients that had not been used to train the model. The testing physicians' personal data were not collected, ensuring maximal data privacy without creating conflicts with employee rights. This study was performed during the period between calendar weeks 11 and 35 in 2023. The evaluation was based on a single Likert scale question ranging from "disturbing" to "excellent."

Structured, Multi-Item Human Evaluation in a Holdout Set of the Training Data

To evaluate the model's performance and potential for real-world application, we conducted a comprehensive study involving a systematic clinician rating of epicrises generated from a set of random, anonymized historical report prompts. This seemed to be the best approach to evaluate the model's performance and to estimate its potential benefits.

The human evaluation was performed using another custom intranet-based Python-programmed web application. Four experienced physicians, with professional experience ranging

XSL•FO
RenderX

from 1 to 20 years, used the evaluation tool to rate the output of the generated epicrises in different metrics that are described later in the study. The physicians were given a guideline on how to evaluate each aspect.

Estimating the Time-Saving Potential

We also developed a method to estimate the time savings for physicians when using AI-generated epicrises. In our multidimensional evaluation tool, we included a section where users copied and pasted the AI-generated epicrisis and revised it to match the quality of a gold-standard human-written version. We measured the time taken for this task. Prior to this, users completed a test to assess their writing speed. These results were then used to calculate the estimated time savings per epicrisis. This approach was intended to account for the possibility that there are mistakes of any kind in the AI-generated text, but just adapting the text would still be a benefit in terms of time consumption compared to writing a completely new epicrisis.

Structured Multi-Item Automated Evaluation Using 2 LLMs

We implemented an automated evaluation technique where the outputs from LLMs were assessed based on several parameters. This evaluation was conducted using the commercially available OpenAI GPT-3.5 Turbo model, alongside Google's open-source Gemma2 model (27B), to provide more nuanced and unbiased results [17,18]. We chose 2 different models to avoid potential bias, as the epicrises being evaluated were generated by an OpenAI model, and using only models from the same ecosystem could introduce bias.

The OpenAI model was accessed through the official API, with the total cost for evaluating all 86 letters amounting to approximately $0.20 USD. These costs were privately funded, and there were no conflicts of interest. The Gemma2 model was accessed via Ollama and installed locally on an Apple MacBook Pro (M3 Processor, 24GB RAM; Apple Inc).

For each evaluation, the medical report, human-written epicrisis (as the gold standard), and AI-generated epicrisis were provided as inputs. The models were prompted to assess whether the AI-generated text was medically accurate, formally correct, provided a time-saving advantage for clinicians, and was overall useful. Responses were requested on a 5-point scale ranging from "I disagree" to "I completely agree," aligning closely with the human evaluation to ensure consistency.

Prompt engineering played a crucial role in obtaining meaningful results. The prompting was achieved using state-of-the-art techniques and libraries, such as LangChain. A total of 86 randomly selected reports were evaluated by both the OpenAI and Gemma2 models, using the same reports as in the human evaluation to enhance repeatability and consistency. Our evaluation prompt was carefully designed to align with the other evaluation metrics, ensuring coherence across all assessments.

Ethical Considerations

The Ethics Committee of the University of Freiburg (24-1370-Anfrage) confirmed that formal ethical approval was not required for this project, as it exclusively involved the use of fully anonymized data. The committee clarified that research using anonymized text data does not constitute human participant research under the current ethical guidelines. All research was conducted in adherence to relevant ethical guidelines and regulations, ensuring compliance with institutional and international standards.

Individual informed consent was not required as this study used only fully anonymized data, with no possibility of identifying individual patients. This approach was validated through a DPIA and approved by our institutional data protection officer.

Data Analysis

Statistical analyses were conducted using Python (SciPy, NumPy, and Pandas). A 1-sample Wilcoxon signed-rank test was used to compare subjective ratings (eg, formal correctness, medical accuracy, and usefulness) against a hypothesized neutral midpoint of 2.5. Paired *t* tests were employed to compare time and length measurements (eg, manual vs AI-corrected writing durations, and character counts). As each AI-corrected epicrisis was compared against the original human-written version for the same medical report, these data were inherently paired. Agreements between human and automated assessments were evaluated descriptively, based on the percentage of categorical rating alignment.

Results

Results Of the 1-Dimensional Clinical Rating

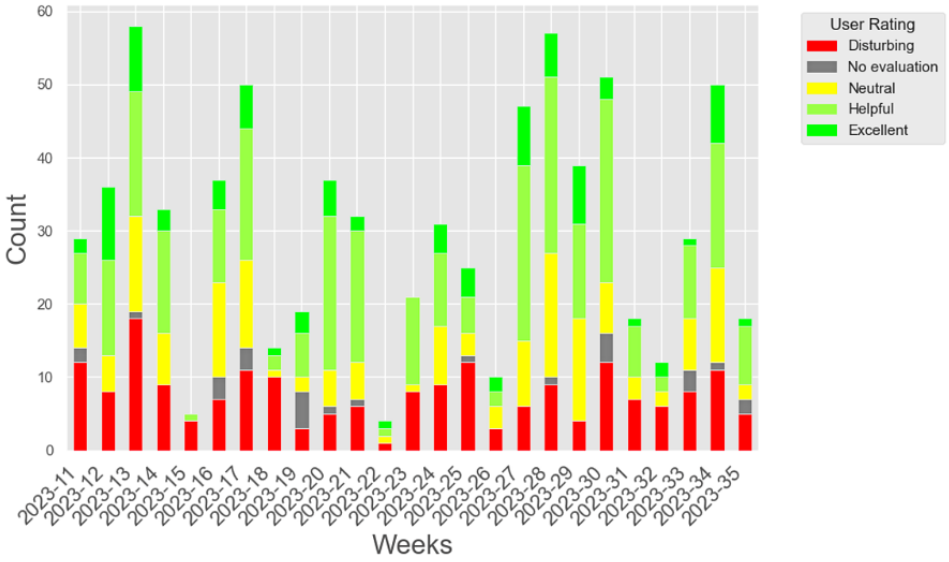
In the real-world evaluation (n=769), the AI-generated epicrises received a mean rating of 1.08 (SD 1.38, coding for helpful) on a scale from –1 (disturbing) to 3 (excellent). Most ratings were positive, with 290 of 769 (37.7%) marked as helpful and 94 of 769 (12.2%) as excellent, while 194 of 769 (25.2%) were rated disturbing, indicating some polarization (Table 1). The results of the real-life clinical rating are shown in Figure 1. Although a 1-sample Wilcoxon signed-rank test did not show a statistically significant deviation from the neutral benchmark of 1 (W=83,038; P=.87), this may be attributable to the broad rating scale and inherent variability in real-world clinical presentations. Given the favorable distribution and positive trend in responses, the results still suggest that the model's output was generally perceived as useful and well-accepted in everyday practice.

XSL•FO
RenderX

Table 1. Distribution of the 1-dimensional human evaluation. The largest percentage was evaluated to be useful.

Rating	Count (N=769), n (%)
Disturbing	194 (25.2)
No evaluation	30 (3.9)
Neutral	161 (20.9)
Helpful	290 (37.7)
Excellent	94 (12.2)

Figure 1. One-dimensional human evaluation recorded during the calendar weeks 11 to 35. The data were obtained over several weeks. Bad ratings are colored in red tones and good ratings in green tones .



Results of the Manual Evaluation

The physician rating was performed by 4 different physicians with clinical experience ranging from 1 year to more than 20 years. The ratings were performed as explained earlier in the Methods section. Table 2 shows the summarized numerical results of the evaluation.

Table 2. Overall physician rating of the AI[a]-generated epicrises. The ratings of all 4 physicians are summarized.

Evaluation metric	1: I disagree (n=107), n (%)	2: I partly agree (n=147), n (%)	3: I mostly agree (n=235), n (%)	4: I completely agree (n=254), n (%)
Formally correct	12 (11.2)	8 (5.2)	48 (20.4)	118 (46.5)
Medically correct	32 (29.9)	48 (32.7)	63 (26.8)	43 (16.9)
Useful	33 (30.8)	47 (32)	67 (28.5)	38 (15)
Advantage in time	30 (28)	44 (29.9)	57 (24.3)	55 (21.7)

[a]AI: artificial intelligence.

The proportion of evaluations representing a helpful AI written epicrisis was in all cases bigger than the proportion representing a rather disturbing or useless one. Figure 2 shows a different approach to visualizing the results grouped for the different metrics to experience a more nuanced view. The evaluations are very balanced for most metrics. In the case of the metric *formally correct*, there was a bigger tendency toward positive evaluations. Wilcoxon signed-rank tests were used to evaluate whether physician ratings differed significantly from a neutral midpoint of 2.5 on the 4-point rating scale. The neutral midpoint of the scale was 2.5, representing a neutral evaluation. The 1-sample Wilcoxon signed-rank test was used to assess if ratings significantly deviated from this midpoint. *Formally correct* was rated significantly higher than neutral (W=1686; P<.001), reflecting strong agreement on structural quality. *Advantage in time* was also rated significantly above the midpoint (W=6564; P<.01), indicating that clinicians perceived a clear time-saving benefit. In contrast, ratings for *medically correct* (W=7456;P=.08) and *usefulness* (W=7652.5; P=.18 did not significantly differ from neutral, suggesting more variable impressions regarding clinical accuracy and usefulness while still showing a quantitative tendency. All statistical results can be found in Table 3.

XSL•FO
RenderX

Figure 2. Physician rating of artificial intelligence (AI)-generated epicrises .

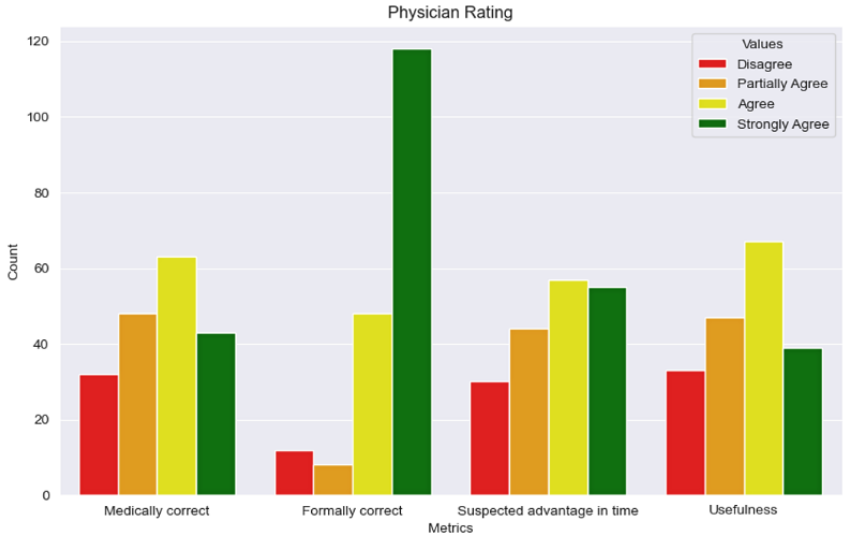


Table 3. Statistical results for the physician ratings.

Metric	Median	*W* statistic	*P* value	Significant (*P*<.05)
Medically correct	3	7456	.08	No
Formally correct	3	1686	<.001	Yes
Useful	3	7652.5	.18	No
Advantage in time	3	6564	.003	Yes

Calculation of Possible Time Advantage

As previously described, we added a way of estimating a possible benefit in time consumption for our study. The time the tester needed to correct the computer-generated epicrisis was measured. The time to write the human-written epicrisis was calculated by dividing the length of the human-written epicrisis by the user's individual writing speed. The average time to rewrite the human epicrisis was 109.52 seconds, while the average time to correct the AI-generated epicrisis was 54.25 seconds, suggesting a time advantage of 56 seconds per epicrisis. The exact results are shown in Table 4.

Table 4. Times and lengths of the epicrises evaluated by 4 different physicians (N=69).

Statistic	Real epicrisis time (seconds)	Corrected epicrisis time (seconds)	AI[a]-generated length (characters)	Real length (characters)	Corrected length (characters)
Mean	109.52	54.25	330.43	501.07	352.22
SD	53.3	63.34	115.42	243.5	135.78
Min	11.35	0.77	125	52	118
25th percentile	73.14	21.94	261	335	255
Median	98.47	44.46	315	451	342
75th percentile	131.86	65.12	400	604	445
Max	261.4	499.86	689	1192	691

[a]AI: artificial intelligence.

What must be considered in this context is that there were significant differences in epicrisis length. To assess whether differences in documentation time were influenced by the length of the epicrises, we compared character counts across versions. AI-generated epicrises were significantly shorter than human-written ones (t_{68}=–6.10; *P*<.001), and the corrected AI versions also remained significantly more concise (t_{68}=–6.65; *P*<.001). These results suggest that the AI system may support more efficient and focused documentation by generating concise summaries of relevant information. This may be particularly beneficial for nonnative speakers, as it helps reduce linguistic complexity while preserving clinical content. The differences are portrayed in a boxplot that can be found in Multimedia Appendix 1. Additionally, this may be due to a person's information advantage when writing, since not all information can be inferred from the letter above. A paired *t* test comparing the time required for generating human-written epicrises versus correcting AI-generated ones revealed a statistically significant

XSL•FO
RenderX

reduction in time with AI assistance (t_{68}=3.39; P=.001). On average, correcting the AI-generated epicrises was notably faster, suggesting a potential efficiency gain in documentation workflows. This result supports the hypothesis that AI-assisted documentation can meaningfully reduce the time burden for clinicians. All exact statistical results are shown in Table 5.

Table 5. Statistical results for time and length comparisons.

Comparison	Mean (SD)	Test type	*t* test (df)	*P* value
Time: human-written vs corrected AI[a]	109.52 (53.3) vs 54.25 (63.3) s	Paired *t* test	3.39 (68)	<.001
Length: human-written vs AI-generated	501 (243.5) vs 330.43 (115.42) characters	Paired *t* test	–6.10 (68)	<.001
Length: human-written vs corrected AI version	501 (243.5) vs 352.22 (135.78) characters	Paired *t* test	–6.65 (68)	<.001

[a]AI: artificial intelligence.

Results of the Automated Evaluation

The automated evaluation showed a tendency for medium values but demonstrated an overall good correspondence with the human evaluation from earlier. Notably, the automated AI evaluation was more hesitant to assign extreme values. The exact results of the evaluations can be seen in Table 6, and they show good overall correspondence. Automated LLM assessments generally aligned with human ratings, with over 52% (356/679) of LLM responses and 66% (489/743) of human ratings falling into the top 2 agreement categories. While this supports overall consistency, direct statistical comparisons (eg, interrater reliability) are not appropriate, as LLMs are not human raters. The evaluation serves as proof of concept for future refinement.

Table 6. AI[a]-generated evaluation of the epicrises. The evaluation was conducted using 2 LLMs[b]. The first was OpenAI's GPT-3.5 Turbo model, accessed through the OpenAI API. The second was a locally run Gemma2 (Google LLC), accessed through Ollama.

Metric	Disagree	Partially agree	Agree	Strongly agree
OpenAI ChatGPT 3.5 Turbo				
Medically correct	0	62	22	1
Formally correct	2	18	64	1
Suspected advantage in time	4	28	44	8
Usefulness	0	63	22	0
Google Gemma2				
Medically correct	0	48	37	0
Formally correct	1	9	74	1
Suspected advantage in time	3	18	53	11
Usefulness	1	66	18	0

[a]AI: artificial intelligence.

[b]LLM: large language model.

Discussion

Principal Findings

Our study had 2 primary objectives. First, we aimed to develop a proof of principle for medical report writing assistance under optimal conditions, providing valuable insights for future developments that could be implemented within EU regulatory frameworks. Second, we sought to assess the potential degree of workload reduction for clinicians.

To achieve the goal of exploring the full potential of current LLM technology in medical report writing, we deliberately chose to use OpenAI's high-resource data centers and powerful models. Although this approach may have limitations for routine usage in the EU due to GDPR constraints, our work may also be directly useful for sites outside the EU without the infrastructure to fine-tune and host LLMs on premises.

By using OpenAI resources, we were able to demonstrate a best-case scenario for LLM-assisted medical report writing. Our results show that under these conditions, there was a significant potential for reducing the administrative burden on clinicians. The real-life clinical ratings, where more than 50% of evaluations found the AI-generated epicrises helpful or excellent, suggest that such a system could substantially decrease the time and effort required for documentation tasks.

The physician ratings and automated evaluations further support this potential for workload reduction. The high scores in categories such as *advantage in time* and *useful* indicate that AI-generated epicrises could save valuable clinician time, allowing for more focus on direct patient care.

The estimation for writing time revealed a potential savings of approximately 56 seconds (nearly 50%) per medical report. Given that physicians typically handle about 20 reports daily, this could lead to a significant reduction in workload. Our study

XSL•FO
RenderX

successfully demonstrated that our approach is a viable solution to achieve these goals, showcasing the potential of LLMs in medical documentation. Additionally, we developed an effective anonymization algorithm that complies with the stringent data protection standards of European law for data transfer to non-EU countries. This accomplishment is significant, as it facilitates collaboration on sensitive patient data while maintaining compliance with GDPR. Our anonymization process, replacing sensitive patient data with placeholders, was crucial in providing the LLM with a high-quality text corpus to work with.

Comparison With Prior Work

Our findings align with and build upon several key studies in this field. There are now several studies that show that the use of LLMs can be helpful in reducing physician workload [6,19].

Van Veen et al [6] demonstrated that adapted LLMs could outperform medical experts in summarizing clinical texts, with their generated summaries often rated as equivalent or superior to human-written ones. This study adopted a broader approach across various medical fields. Heilmeyer et al [8] explored the use of open-source LLMs hosted on premises for a similar task, focusing on the viability of local, privacy-preserving solutions. Other approaches that could also be combined with our work in a helpful context include that of Kernberg et al [20], who used LLMs to create medical notes from audio files. Singh et al [21] successfully used LLMs to create discharge summaries and operative notes. We believe this demonstration suggests that our approach could potentially be applied in a broader context while working with the same techniques.

Limitations

We recognize that the time-saving potential is merely an estimate, and actual results may vary in practice, which we could not evaluate at this stage. A standardized assessment of the time-saving potential is methodically challenging to quantify.

In addition, it is crucial to note that these results represent an ideal scenario, while real-world implementation may present additional challenges. This might include retraining the fine-tuned models or developing models that are capable of online training to adapt to changing guidelines and state-of-the-art medical treatment. The development of models with reinforcement learning may be a good approach and should be considered in further research.

While our anonymization process was conducted to fulfill the high data privacy standards according to the laws in Germany and Europe, it must still be considered that complete anonymization is difficult to achieve. There are cases of very rare diseases, such as corneal or retinal dystrophies, or other individual conditions that could potentially be identified if supplementary data were available.

Our training corpus has limitations in terms of size and domain diversity, which may affect the model's adaptability. We observed that the model performs best on very common, standardized cases in our daily clinical routine, such as pterygium, keratoconus, and age-related macular degeneration. On the other hand, the model sometimes struggles with more complex cases involving individualized treatment plans. This has motivated us to continue in-house model development to improve the training corpus and retrain a new model in the future. It is also worth noting that the model may perform differently when applied to other medical disciplines. Ophthalmology, as a specialized field, may present challenges for generally trained models, while areas like internal medicine, which benefit from more extensive training data, may see better results.

While our approach is relatively straightforward to implement, significant challenges remain. Many hospitals do not have structured access to medical reports and still use unstructured formats like Microsoft Word, complicating automated data processing. Although our model setup is simple, delivering data to APIs such as OpenAI in a structured JSON format is crucial. For clinics that can achieve this, the training process becomes easier compared to developing in-house hardware and software for model fine-tuning. However, considerable domain knowledge is still required for effective implementation, which may be lacking in many hospitals.

Evaluating the performance of language models in medical contexts is inherently challenging, which is why we implemented a variety of metrics to assess our model's performance. We aimed to create a set of possibilities that can be used to evaluate the performance of LLMs. This mainly relied on human ratings in different contexts. We also included an AI auto-evaluation as a proof of concept that could be applied in other contexts with bigger sample sizes that are difficult to evaluate manually. Since the human and the AI auto evaluation showed good alignment, we believe this approach can be used for similar tasks in the future.

Several factors must be considered when evaluating our project and identifying areas for future research. The rapid advancement of generative AI requires ongoing reevaluation.

One limitation in this context is the fact that this study was conducted using the OpenAI Curie model, which has since been deprecated. While the specific model is no longer state-of-the-art, the primary contribution of our work lies in the demonstration of a feasible end-to-end pipeline. The methodologies and principles established here are directly transferable to newer, more powerful models, and we are actively conducting ongoing research with these models.

Additionally, other groups within our hospital have achieved success using on-premises solutions for this task. However, the training and fine-tuning processes remain challenging because of the demand for highly qualified personnel as well as hardware requirements. Our study did not perform a direct head-to-head comparison, and the hypothesis that commercial models outperform local ones remains a topic for future empirical validation.

Conclusion

In conclusion, our study demonstrates the significant potential of LLMs in medical report generation under optimal conditions, provides a quantitative assessment of possible workload reduction for clinicians, and highlights the challenges of implementing such systems within the EU's regulatory framework. The proof of principle established here sets a

XSL•FO
RenderX

benchmark for what is achievable with the current LLM technology. Future research should focus on developing equally powerful solutions that can be deployed on premises or within EU-based cloud services, ensuring both high performance and full compliance with the GDPR. We hope that this study will guide the development of EU-compliant solutions capable of matching or exceeding this performance level while significantly reducing the administrative burden on health care professionals—allowing physicians to focus rather on patient care than on medical documentation. Our work represents one step toward achieving this goal [3].

Acknowledgments

We would like to thank all physicians who participated in the evaluation process for their time and insights. We are grateful to the Department of Ophthalmology of the University Hospital Freiburg for providing the necessary infrastructure. We appreciate the help of the data protection team as well as the local ethics committee for their assistance in ensuring that our study was GDPR-compliant. This study received no additional funding from external sources. The costs for the fine-tuning process and all hardware requirements were paid by the institution. Generative AI and AI-assisted technologies (DeepL, GPT-4o, and Gemini) were used to make linguistic improvements during the writing process. The authors carefully checked the text suggestions for accuracy and made corrections where necessary. They take full responsibility for the publication's content. We acknowledge support by the Open Access Publication Fund of the University of Freiburg.

Data Availability

The data sets used and/or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

SA and DB were responsible for data acquisition, training, fine-tuning, and implementation of the model. SA and DB also created the evaluation software. DB and SA were responsible for integrating the model into clinical production. AR, QVN, LS, and SA performed the model evaluation manually. TR and FH were responsible for overseeing the project. SA was the main manuscript author. Extensive proofreading was performed by LS and FH. All authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Calculated time to write the human-written epicrises with the respective writing speed of the tester and lengths of the epicrises in characters.
[PNG File , 37 KB - ijmr_v14i1e72894_app1.png]

References

1. Robertson SL, Robinson MD, Reid A. Electronic health record effects on work-life balance and burnout within the I3 population collaborative. J Grad Med Educ 2017 Aug;9(4):479-484. [doi: 10.4300/jgme-d-16-00123.1]
2. Overhage JM, McCallie D. Physician time spent using the electronic health record during outpatient encounters: a descriptive study. Ann Intern Med 2020 Feb 04;172(3):169-174 [FREE Full text] [doi: 10.7326/M18-3684] [Medline: 31931523]
3. Yu P, Xu H, Hu X, Deng C. Leveraging generative AI and large language models: a comprehensive roadmap for healthcare integration. Healthcare (Basel) 2023 Oct 20;11(20):2776 [FREE Full text] [doi: 10.3390/healthcare11202776] [Medline: 37893850]
4. Karabacak M, Margetis K. Embracing large language models for medical applications: opportunities and challenges. Cureus 2023 May;15(5):e39305 [FREE Full text] [doi: 10.7759/cureus.39305] [Medline: 37378099]
5. Haghighi T, Gholami S, Sokol JT, Kishnani E, Ahsaniyan A, Rahmanian H, et al. EYE-Llama, an in-domain large language model for ophthalmology. iScience 2025 Jul 18;28(7):112984 [FREE Full text] [doi: 10.1016/j.isci.2025.112984] [Medline: 40697409]
6. Van Veen D, Van Uden C, Blankemeier L, Delbrouck J, Aali A, Bluethgen C, et al. Adapted large language models can outperform medical experts in clinical text summarization. Nat Med 2024 Apr 27;30(4):1134-1142. [doi: 10.1038/s41591-024-02855-5] [Medline: 38413730]
7. Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, et al. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions. ACM Trans Inf Syst 2025 Jan 24;43(2):1-55. [doi: 10.1145/3703155]
8. Heilmeyer F, Böhringer D, Reinhard T, Arens S, Lyssenko L, Haverkamp C. Viability of open large language models for clinical documentation in German health care: real-world model evaluation study. JMIR Med Inform 2024 Aug 28;12:e59617-e59617. [doi: 10.2196/59617] [Medline: 39195570]

XSL•FO
RenderX

9. NVIDIA Hopper GPUs expand reach as demand for AI grows. NVIDIA News. URL: https://nvidianews.nvidia.com/news/nvidia-hopper-gpus-expand-reach-as-demand-for-ai-grows [accessed 2025-05-15]
10. Bradford L, Aboy M, Liddell K. International transfers of health data between the EU and USA: a sector-specific approach for the USA to ensure an 'adequate' level of protection. J Law Biosci 2020;7(1):lsaa055. [doi: 10.1093/jlb/lsaa055] [Medline: 34221424]
11. EU Regulatory Data Protection: International data transfer rules for non-personal data. DLA Piper. URL: https://tinyurl.com/mvpa2bhc [accessed 2024-09-08]
12. Hallinan D, Bernier A, Cambon-Thomsen A, Crawley FP, Dimitrova D, Medeiros CB, et al. International transfers of personal data for health research following Schrems II: a problem in need of a solution. Eur J Hum Genet 2021 Oct 06;29(10):1502-1509 [FREE Full text] [doi: 10.1038/s41431-021-00893-y] [Medline: 33953344]
13. International transfers -. European Data Protection Supervisor. URL: https://www.edps.europa.eu/data-protection/data-protection/reference-library/international-transfers_en [accessed 2024-09-08]
14. Liss J, Peloquin D, Barnes M, Bierer BE. Demystifying for the cross-border transfer of clinical research data. J Law Biosci 2021;8(2):lsab032 [FREE Full text] [doi: 10.1093/jlb/lsab032] [Medline: 34703612]
15. Wiest I, Leßmann ME, Wolf F, Ferber D, Treeck M, Zhu J, et al. Deidentifying medical documents with local, privacy-preserving large language models: the LLM-anonymizer. N Engl J Med AI 2025 Mar 27;2(4):2024 (forthcoming). [doi: 10.1056/AIdbp2400537]
16. OpenAI Platform. URL: https://platform.openai.com [accessed 2024-09-08]
17. Ollama Gemma2. URL: https://ollama.com/library/gemma2 [accessed 2024-09-08]
18. Gemma 2 is now available to researchers and developers. Google Blog. URL: https://blog.google/technology/developers/google-gemma-2/ [accessed 2024-09-08]
19. Chen C, Liao CT, Tung YC, Liu CF. Enhancing healthcare efficiency: integrating ChatGPT in nursing documentation. Stud Health Technol Inform 2024 Aug 22;316:851-852. [doi: 10.3233/SHTI240545] [Medline: 39176926]
20. Kernberg A, Gold JA, Mohan V. Using ChatGPT-4 to create structured medical notes from audio recordings of physician-patient encounters: comparative study. J Med Internet Res 2024 Apr 22;26:e54419 [FREE Full text] [doi: 10.2196/54419] [Medline: 38648636]
21. Singh S, Djalilian A, Ali MJ. ChatGPT and ophthalmology: exploring its potential with discharge summaries and operative notes. Semin Ophthalmol 2023 Jul 03;38(5):503-507. [doi: 10.1080/08820538.2023.2209166] [Medline: 37133418]

Abbreviations

AI: artificial intelligence
BCR: Binding Corporate Rule
DPIA: Data Protection Impact Assessment
EU: European Union
GDPR: General Data Protection Regulation
GPU: graphics processing unit
HIPAA: Health Insurance Portability and Accountability Act
LLM: large language model
SCC: Standard Contractual Clause

Edited by N Cahill, T de Azevedo Cardoso; submitted 21.02.25; peer-reviewed by LH Yao, M Wang; comments to author 12.05.25; revised version received 21.05.25; accepted 30.09.25; published 26.11.25.

Please cite as:
Arens S, Ngo QV, Richling A, Stürzbecher L, Böhringer D, Reinhard T, Heilmeyer F
Streamlining Ophthalmic Documentation With Anonymized, Fine-Tuned Language Models: Feasibility Study
Interact J Med Res 2025;14:e72894
URL: https://www.i-jmr.org/2025/1/e72894
doi:10.2196/72894
PMID:

©Sebastian Arens, Quang Vinh Ngo, Anna Richling, Lucas Stürzbecher, Daniel Böhringer, Thomas Reinhard, Felix Heilmeyer. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.11.2025. 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

XSL•FO
RenderX

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

Current Status and Future Directions of Ferroptosis Research in Breast Cancer: Bibliometric Analysis

Jia-Yuan Luo[1*], MD; Yu-Long Deng[2*], BMed; Shang-Yi Lu[3], BMed; Si-Yan Chen[4], BMed; Rong-Quan He[2], MD, PhD; Di-Yuan Qin[5], BS; Bang-Teng Chi[2], BMed; Gang Chen[1], MD, PhD; Xia Yang[1], MD, PhD; Wei Peng[2], MD, PhD

[1]Department of Pathology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[2]Department of Oncology, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[3]Department of Hepatological and Gland Surgery, Wuzhou Gongren Hospital/The Seventh Affiliated Hospital of Guangxi Medical University, Wuzhou, China

[4]Day Chemotherapy Center, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

[5]Department of Computer Science and Technology, School of Computer and Electronic Information, Guangxi University, Nanning, China

[*]these authors contributed equally

Corresponding Author:
Wei Peng, MD, PhD
Department of Oncology, The First Affiliated Hospital of Guangxi Medical University
No.6 Shuangyong Road
Nanning, 530021
China
Phone: 86 15977779804
Email: pengwei@stu.gxmu.edu.cn

Abstract

Background: Ferroptosis, as a novel modality of cell death, holds significant potential in elucidating the pathogenesis and advancing therapeutic strategies for breast cancer.

Objective: This study aims to comprehensively analyze current ferroptosis research and future trends, guiding breast cancer research advancements and innovative treatment strategies.

Methods: This research used the R package Bibliometrix (Department of Economic and Statistical Sciences at the University of Naples Federico II), VOSviewer (Centre for Science and Technology Studies at Leiden University), and CiteSpace (Drexel University's College of Information Science and Technology), to conduct a bibliometric analysis of 387 papers on breast cancer and ferroptosis from the Web of Science Core Collection. The analysis covers authors, institutions, journals, countries or regions, publication volumes, citations, and keywords.

Results: The number of publications related to this field has surged annually, with China and the United States collaborating closely and leading in output. Sun Yat-sen University stands out among the institutions, while the journal *Frontiers in Oncology* and the author Efferth T contribute significantly to the field. Highly cited papers within the domain primarily focus on the induction of ferroptosis, protein regulation, and comparisons with other modes of cell death, providing a foundation for breast cancer treatment. Keyword analysis highlights the maturity of glutathione peroxidase 4-related research, with breast cancer subtypes emerging as motor themes and the tumor microenvironment, immunotherapy, and prognostic models identified as basic themes. Furthermore, the application of nanoparticles serves as an additional complement to the basic themes.

Conclusions: The current research status in the field of ferroptosis and breast cancer primarily focuses on the exploration of relevant theoretical mechanisms, whereas future trends and mechanisms emphasize the investigation of therapeutic strategies, particularly the clinical application of immunotherapy related to the tumor microenvironment. Nanotherapy has demonstrated significant clinical potential in this domain. Future research directions should deepen the exploration in this field and accelerate the clinical translation of research findings to provide new insights and directions for the innovation and development of breast cancer treatment strategies.

(Interact J Med Res 2025;14:e66286) doi:10.2196/66286

XSL•FO
RenderX

KEYWORDS

breast cancer; ferroptosis; bibliometric; malignancy; cancer studies; treatment; bibliometric analysis; VOSviewer; China; United States; breast carcinoma; mammary cancer; strategy; trends; bibliography; review; disparities; forecast; treatment strategies; advancements

Introduction

Based on 2022 International Agency for Research on Cancer data, the global burden of breast cancer is alarming, with 2.3 million new cases, and over 665,000 deaths annually. Breast cancer constitutes 11.6% of new cancers and 6.9% of cancer deaths, underscoring the need for urgent research and intervention [1]. It has been projected that by 2040, due to global population growth and aging, the cancer burden will further increase, with breast cancer experiencing particularly notable growth. The number of new cases and deaths are expected to rise to over three million and one million, respectively [2]. As the most prevalent cancer worldwide, surpassing lung cancer, breast cancer poses a significant threat to human health. Despite the considerable progress achieved through sustained research in breast cancer diagnosis, treatment, and prevention, the intricate etiology of this malignancy remains widely undeciphered. Consequently, there is an urgent need to delve deeper into the complex mechanisms that underlie its development, with the aim of developing more precise and efficacious strategies for its management [3]. There is considerable variation in treatment response among individuals with this disease, and the prognosis and treatment outcomes for individuals diagnosed with advanced breast cancer are notably unfavorable [4]. Exploring new mechanisms can enhance our understanding of the biological behavior of breast cancer and improve treatment sensitivity.

Ferroptosis has added a new theoretical basis to the study of the molecular mechanisms underlying breast cancer. Since the introduction of this concept, the popularity of research in this field has rapidly increased, showing an exponential growth trend [5]. Ferroptosis is a recently discovered cell death mechanism mediated by iron-dependent phospholipid peroxidation and is tightly regulated by various cellular metabolic networks, including redox balance, iron handling, mitochondrial function, and amino acid, lipid, and glucose metabolism. This intricate regulation highlights its significance in breast cancer research [6]. Studies on breast cancer have revealed that the regulatory network of ferroptosis is sophisticated and complex, involving the synergistic effects of multiple signaling pathways and transcription factors. This regulatory network may modulate the malignant biological behavior of breast cancer cells via autophagy [7], metabolism [8], immune regulation [9], and other procedures, opening novel strategic directions for breast cancer treatment [5,6,10]. However, the literature concerning breast cancer and ferroptosis currently appears highly fragmented, and a systematic bibliometric analysis to consolidate this information is lacking, making it difficult to accurately summarize the research status and future direction of this field.

In response, this study aims to use bibliometric methods to conduct a comprehensive and objective analysis of the current status and future trends in breast cancer and ferroptosis research. This will entail a rigorous examination of pertinent indicators, including author distribution, research institutions, journals, country or region contributions, scientific publication output, highly cited papers, and keyword analysis. The ultimate objective of this study is to provide an in-depth revelation of the research hotspots and potential development trends in this field, thereby offering guidance for future research endeavors in exploring the mechanisms and therapeutic strategies related to ferroptosis in breast cancer.

Methods

Data Sources and Search Strategy

This study used data from two core collections within the Web of Science Core Collection (WOSCC): the Social Sciences Citation Index and the Science Citation Index Expanded. An advanced search was conducted using the search query (TS=(“Breast Tumor*” OR “Breast Cancer” OR “Mammary Cancer” OR “Breast Carcinoma” OR “Carcinomas, Breast”) AND TS = (“Ferroptosis”)), yielding a total of 624 papers. The search was further refined to include papers published between January 1, 2016, and February 1, 2024. Paper types were limited to original research papers, excluding reviews, conference abstracts, news items, advertisements, and non-English literature that did not meet the inclusion criteria. All data records in this study consisted of complete records and cited references and were exported in plain text format. To avoid errors and ambiguities due to database updates, all searches and data exports were completed on February 18, 2024. Additionally, all data in this study were obtained from open-source public databases; thus, no ethical issues were involved. The search and screening processes were independently conducted by JYL and YLD. In the case of any disagreements during the screening process, the final decision was reached after a discussion with the corresponding authors WP and XY.

Inclusion and Exclusion Criteria

The inclusion criteria were (1) published journal papers; (2) papers with “breast cancer” and “ferroptosis” in the title, indicating a primary focus on breast cancer and ferroptosis; (3) papers not mentioning “ferroptosis” in the title but including ferroptosis-related research on breast cancer in the abstract; and (4) papers containing actual research on the relationship between breast cancer and ferroptosis within the full text.

The exclusion criteria were (1) nonliterature records, such as reviews, conference reports, briefings, news articles, editorial materials, letters, and advertisements; (2) non-English literature, (3) duplicate or data-deficient papers; and (4) withdrawn papers.

Data Standardization and Visualization Methods in Bibliometrics

To ensure data accuracy and consistency, rigorous data standardization processing was applied to the raw data in this study. First, the study uniformly standardized the names of

XSL•FO
RenderX

authors and institutions. Second, controversies over the attribution of country or region names were carefully resolved to ensure that the geographical attributes of the data were clear and unambiguous. In addition, synonymous keywords with the same meaning were merged, and keywords that were not directly related to the specific research content were deleted.

The study used the following tools for visualization and bibliometric analysis: the R package Bibliometrix (version 4.3.2; Department of Economic and Statistical Sciences at the University of Naples Federico II), VOSviewer (version 1.6.19; Centre for Science and Technology Studies [CWTS] at Leiden University), and CiteSpace (version 6.1.R6; Drexel University's College of Information Science and Technology). The specific content included visual analyses of annual scientific publication output and citation trends, authors, institutions, journals, countries or regions, highly cited papers, and keywords. Notably, in VOSviewer, we applied "full counting" for normalized data processing. For cluster analysis of authors, institutions, and countries or regions, a minimum threshold of three publications and citations was set, and the clustering map of the top 100 keywords was further presented. Additionally, the Walktrap method was used in thematic map analysis for more precise topic delineation, while other analyses adhered to default software thresholds. By using these visualization methods, this study provided researchers in related fields with an intuitive and clear understanding of the research landscape.

Ethical Considerations

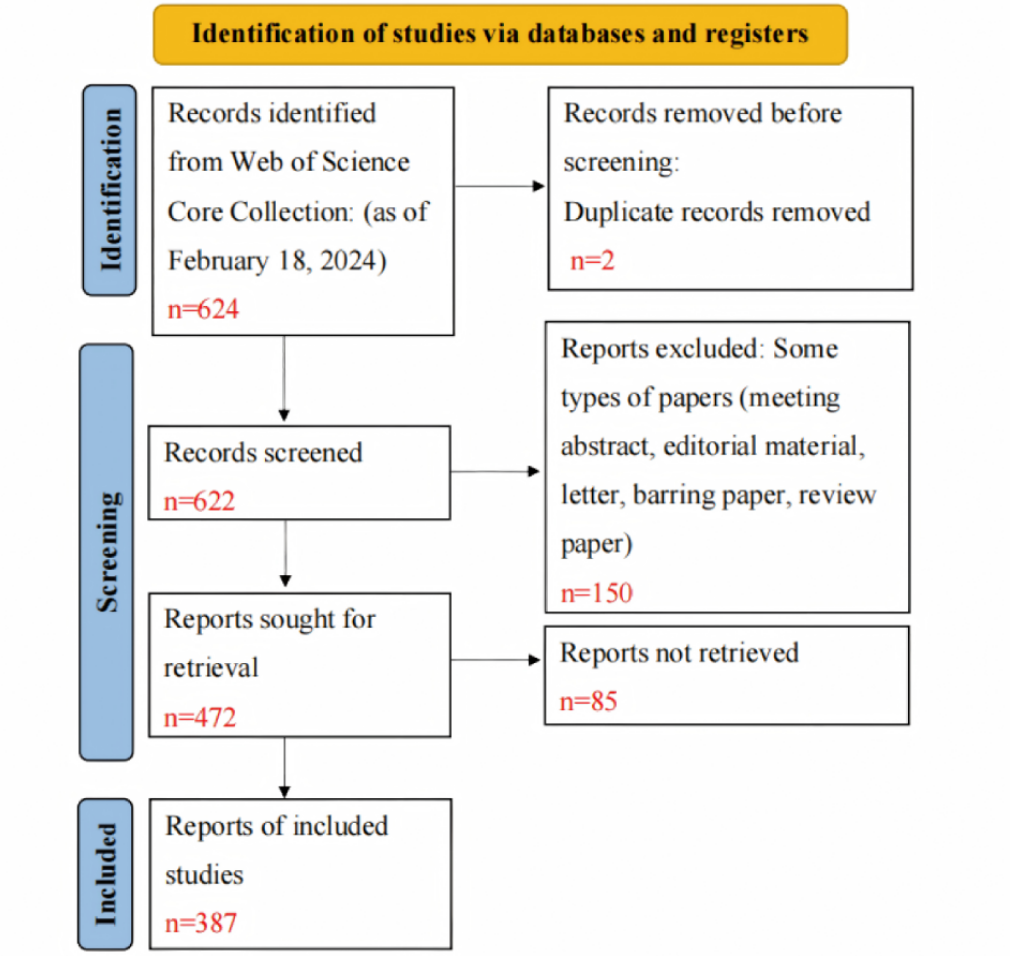
No animal or human studies were carried out by the authors for this paper.

Results

Key Information

This study included a total of 387 academic papers published by 2494 authors in 192 academic journals as the data source (the detailed data collection process is illustrated in Figure 1). On average, each paper had 8.26 coauthors, with only two papers authored by a single individual. The proportion of international collaborations was 16.54%. A total of 840 keywords and 15,107 cited references were identified. It is noteworthy that the average publication age of papers in this field was only 2.02 years, while the average number of citations per paper was as high as 27.22. These data demonstrate that this topic is rapidly emerging as a new research focus (Multimedia Appendix 1).

Figure 1. Flowchart of literature data included in this study.



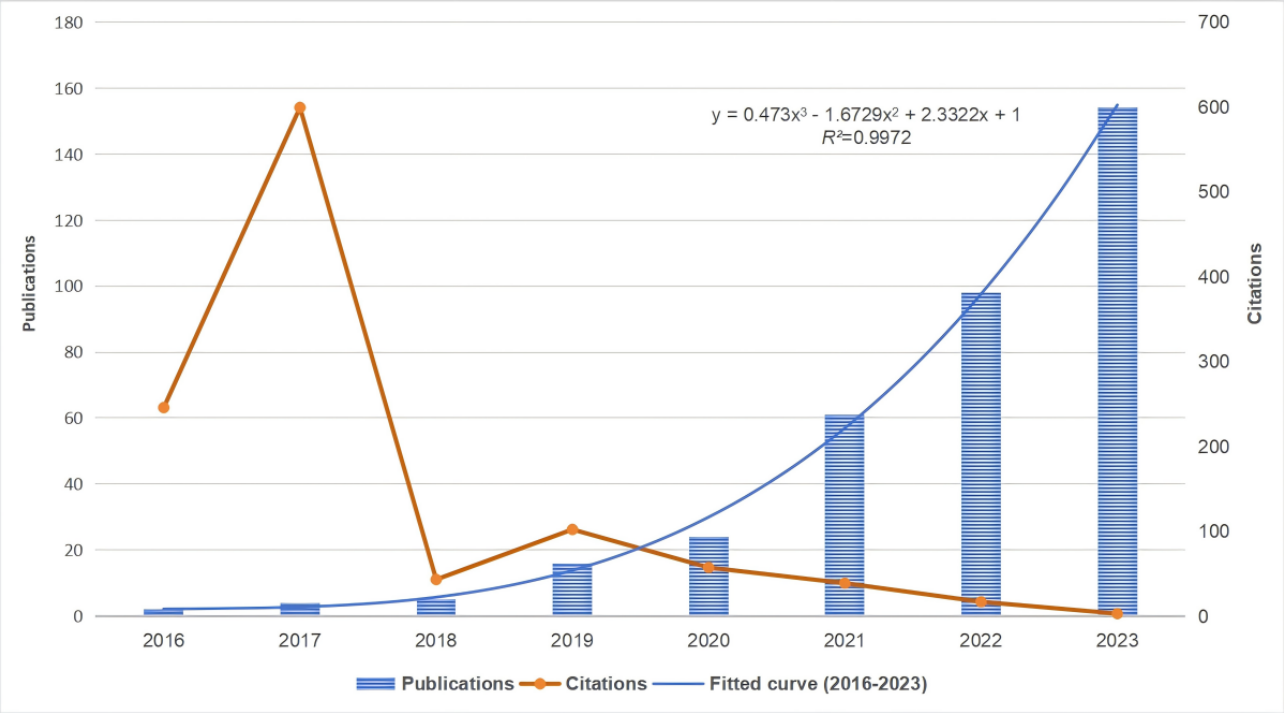
XSL•FO
RenderX

Analysis of Scientific Output in Breast Cancer– and Ferroptosis-Related Research

Annual Analysis of Scientific Publications

Due to the limited availability of relevant literature data published in 2024 up to the retrieval cutoff date of February 1, this study collected complete annual data spanning from 2016 to 2023 (n=364) from a total of 387 papers. A cubic function was fitted to the data to analyze the trend in publication volume (R^2= 0.9972; Figure 2). Specifically, the bar chart indicates that the first paper in this field was published in 2016 (n=1), with 2023 marking the peak year of annual publications (n=154). The line graph represents the average annual citation frequency of relevant literature from 2016 to 2023, revealing a peak of 599 citations in 2017, followed by an overall decreasing trend in subsequent years.

Figure 2. Visualization of annual publication and citation trends of papers (2016-2023). Note that the literature data for 2024 is only counted up to February 1, 2024, with 23 publications and 0 citations.

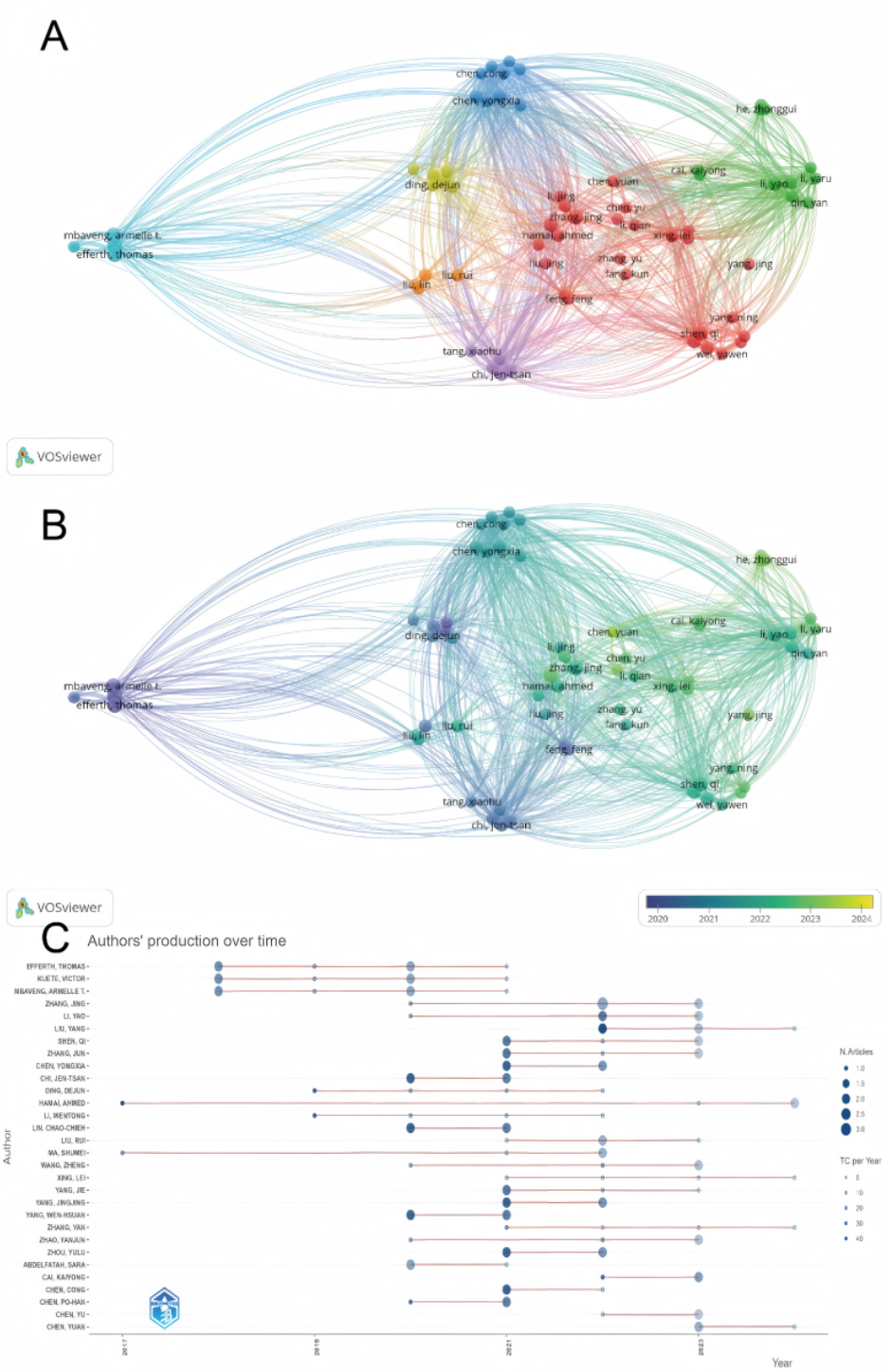


Analysis of Scientific Publications by Prolific Authors

Through an analysis of the scientific publication output of prolific authors, the study found that 22 distinguished authors had published more than two papers in this research field, with no significant differences in the publication volume among the author groups in the field of ferroptosis and breast cancer. Notably, a minority of authors, represented by Efferth et al, had a significantly higher publication output, reaching six papers. Closely following them were authors from the Shen and Zhang groups, who have also contributed significantly to the lactoferrin research landscape, with a noteworthy publication count of five papers each, occupying important positions in the field of breast cancer and ferroptosis research (Figure 3A). It is noteworthy that, excluding the incomplete data from 2024, as shown in Figures 3B and 3C, an increasing number of new distinguished authors have emerged in recent years, with their publications mainly concentrated after 2020.

XSL•FO
RenderX

Figure 3. Visualization charts of authors' publication volume and publication trends. (A) Network visualization map of authors' publication volume. (B) Stacked visualization map of authors' publication volume. (C) Heatmap of authors' publication volume.

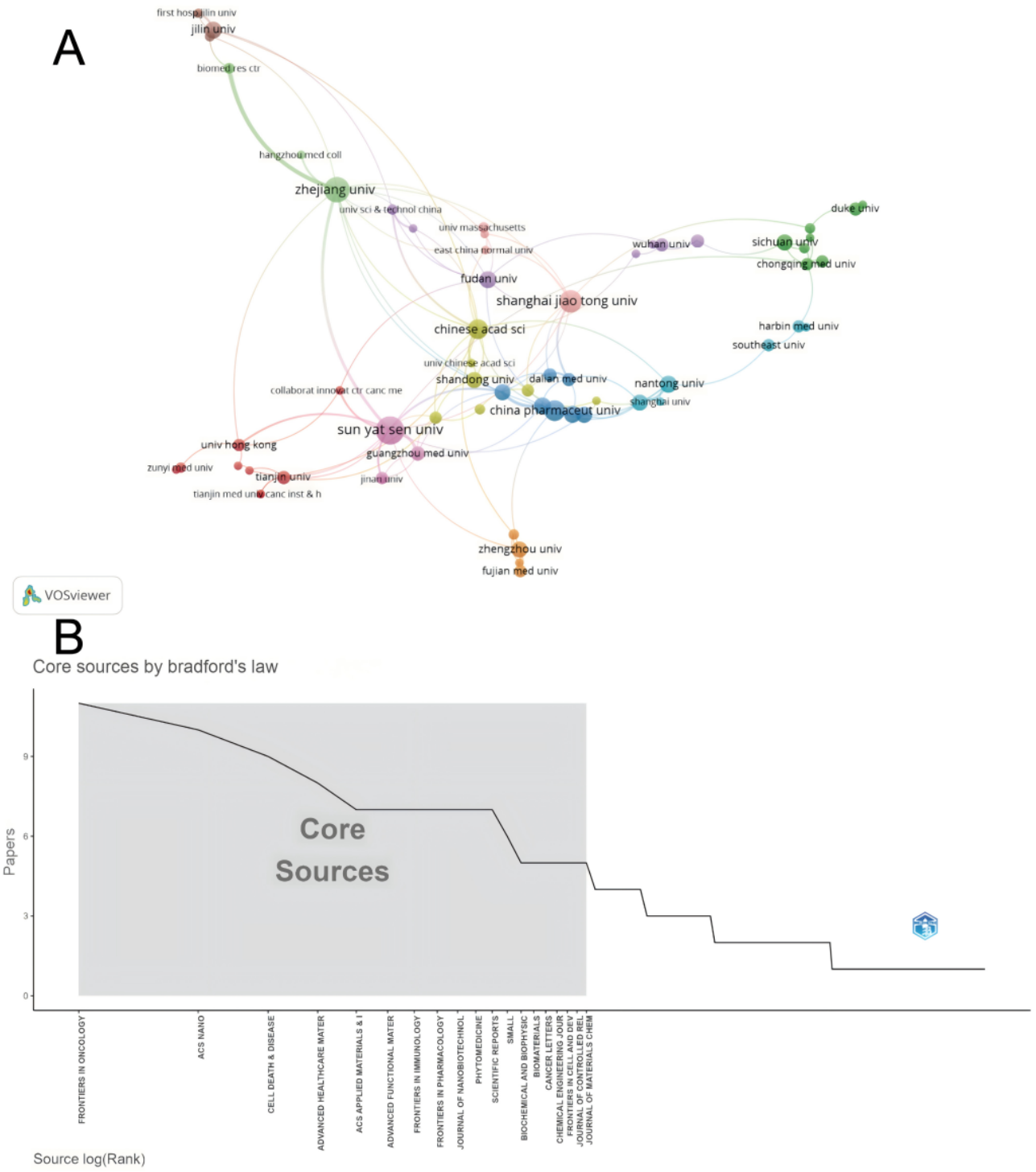


Analysis of Scientific Publications by Prolific Institutions

By analyzing the bibliometric data related to the research fields of breast cancer and ferroptosis, this study identified a group of institutions with significant scientific output in this area (Figure 4A). Among them, Sun Yat-sen University (n=23 publications) and Zhejiang University (n=19 publications) lead the list of institutions with publications related to breast cancer and ferroptosis. Apart from these two universities, other renowned domestic institutions, such as Shanghai Jiao Tong University (n=15 publications), China Pharmaceutical University (n=13 publications), and the Chinese Academy of Sciences (n=12 publications) have also performed exceptionally well in this field. It is also worth mentioning that non-Chinese institutions, such as Duke University (n=6 publications) and Stanford University (n=3 publications), also possess research capabilities in this area.

XSL•FO
RenderX

Figure 4. Visualization maps of institutions' publication volume and identification of core journals. (A) Network visualization map of scientific institutions' publication volume. (B) Identification of journals that have made significant contributions to the field, based on Bradford's Law.

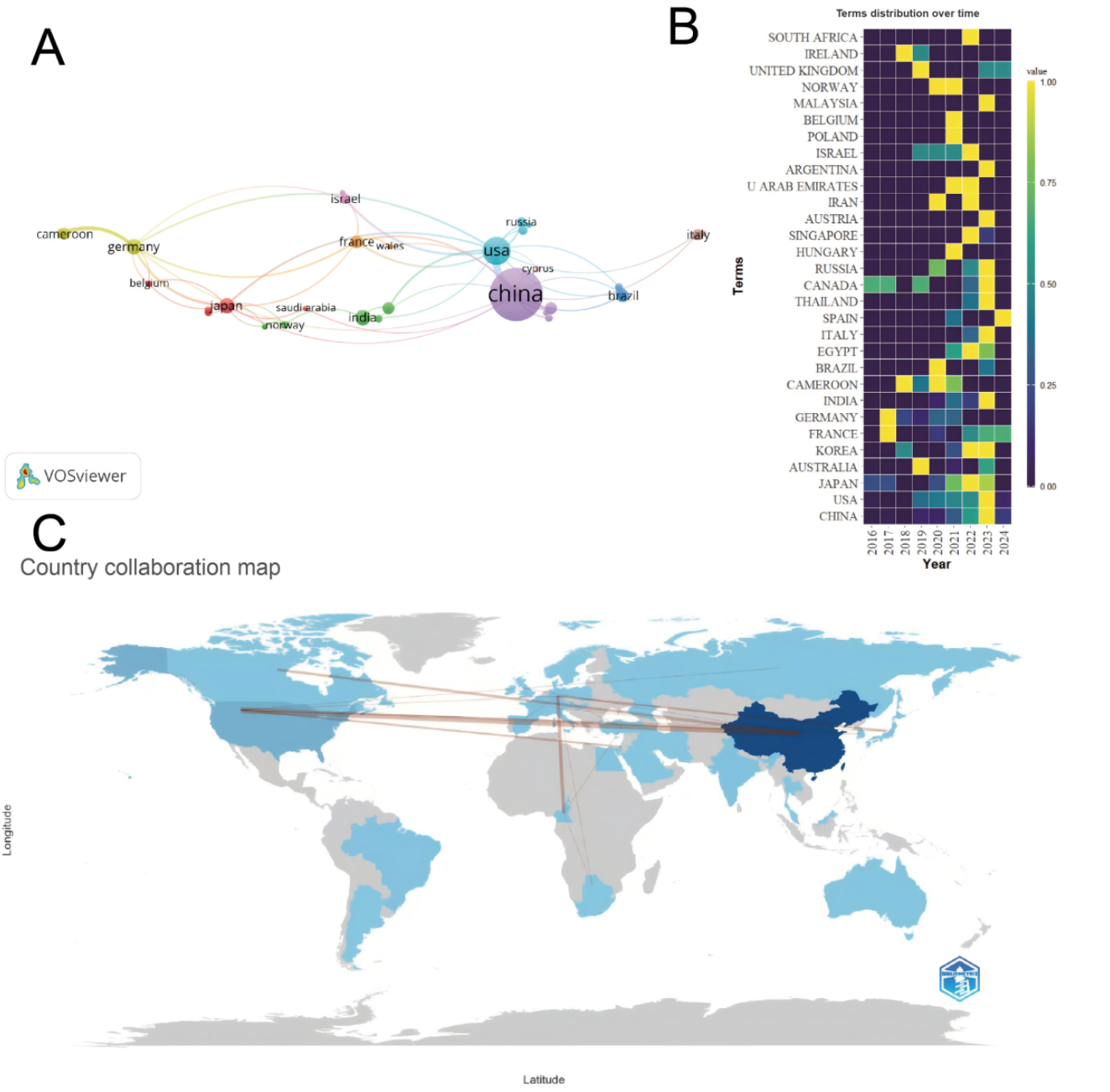


Analysis of Scientific Publications by Prolific Countries

The study found that China (n=287) and the United States (n=52) occupy a dominant position in terms of paper output. Furthermore, countries such as Germany (n=11), India (n=11), Japan (n=11), France (n=7), Brazil (n=7), and Cameroon (n=6) have achieved certain research outcomes in the study of breast cancer and ferroptosis (Figure 5A). Furthermore, as depicted in Figure 5B, China experienced a pronounced acceleration in the volume of scientific publications focused on this research area around 2023, exhibiting a remarkable growth pattern. Moreover, the establishment of a close collaborative partnership between China and the United States emphasizes the inherently global nature of research endeavors in this domain (Figure 5C).

XSL•FO
RenderX

Figure 5. Global publication landscape and collaboration in breast cancer and ferroptosis research. (A) Network visualization map of country publication volume. (B) Heatmap of country publication trends. (C) Map of country collaboration relationships.



Analysis of Citation Volume in Breast Cancer– and Ferroptosis-Related Research

Analysis of Frequently Cited Papers

As shown in the data presented in Multimedia Appendix 2, this study found that the highest citation count was from the paper "Ferroptosis is induced following siramesine and lapatinib treatment of breast cancer cells," published in the journal *Cell Death & Disease*, with 38 citations. From the perspective of research focus and content, these citations primarily concentrate on the following aspects: first, the mechanisms and impacts of ferroptosis induction in breast cancer cells by distinctive drugs or therapeutic methods, as exemplified by papers 1, 2, 3, 4, 6, and 10; second, the exploration of the web-based mechanisms between related protein modulation and ferroptosis variations in breast cancer, as demonstrated in papers 2, 5, 7, 8, and 9; and third, the contrasts and interrelationships between ferroptosis and other modes of cell death in breast cancer cells, as discussed in paper 9.

Analysis of Frequently Cited Authors

The contributions and recognition of different authors can be reflected by the total citation count of their papers. According to the author citation analysis in Multimedia Appendix 3, Chen Y ranks at the top among all authors, with a total citation count of 64. Following closely is Gibson SB, with a total citation count of 51. Additionally, papers by Ding DJ, Li WT, Gai CC, Henson EE, Li ZHR, and Ma S have all been cited 38 times or more.

XSL•FO
RenderX

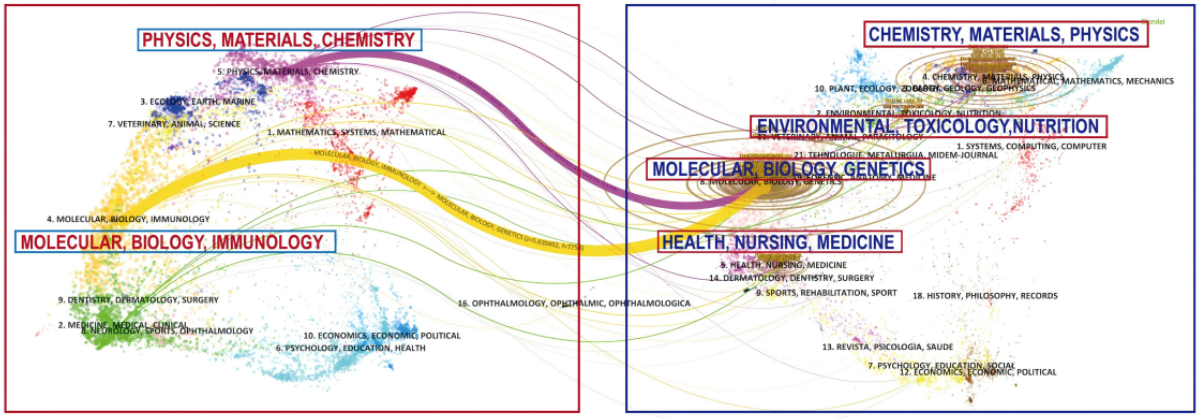
Analysis of Frequently Cited Countries

There are notable differences in the citations of research outcomes from different countries. According to the results shown in Multimedia Appendix 4, China ranks at the top in terms of total citations, with 4877, but the average number of citations per paper is 17.40. Conversely, Germany boasts the highest average citations per paper, reaching an impressive 489.20 citations. The United Kingdom, France, and Canada follow closely with average citations of 187.00, 118.70, and 108.80 per paper, respectively. It is noteworthy that the United States while ranking third after China and Germany in total citations with 1658, records an average of 50.20 citations per paper.

Overlay Analysis of High-Frequency Citing and Cited Journals

In this study, CiteSpace software was used to construct a dual-layer overlay map of journals, visually representing the intricate relationships between citing journals (located on the left) and cited journals (located on the right) within the research domain of breast cancer and ferroptosis (Figure 6). The analysis results revealed that the group of citing journals represented by "Molecular, Biology, Immunology" and the core group of cited journals centered around "Molecular, Biology, Genetics" constitute the pivotal nodes in the knowledge flow path at the journal level (*z* score=5.639492; *f*=7758). Closely following this, there is a strong connection between the citing journal group of "Physics, Materials, Chemistry" and the cited journal group of "Molecular, Biology, Genetics" (*z* score= 2.1142766; *f*=3189).

Figure 6. Dual-map overlay of journals illustrating the citation relationships between citing journals on the left and cited journals on the right. The width of the connecting lines represents the strength of the citation relationships.



Cluster Analysis of High-Frequency Keywords and Main Evolution Directions in Breast Cancer– and Ferroptosis-Related Research

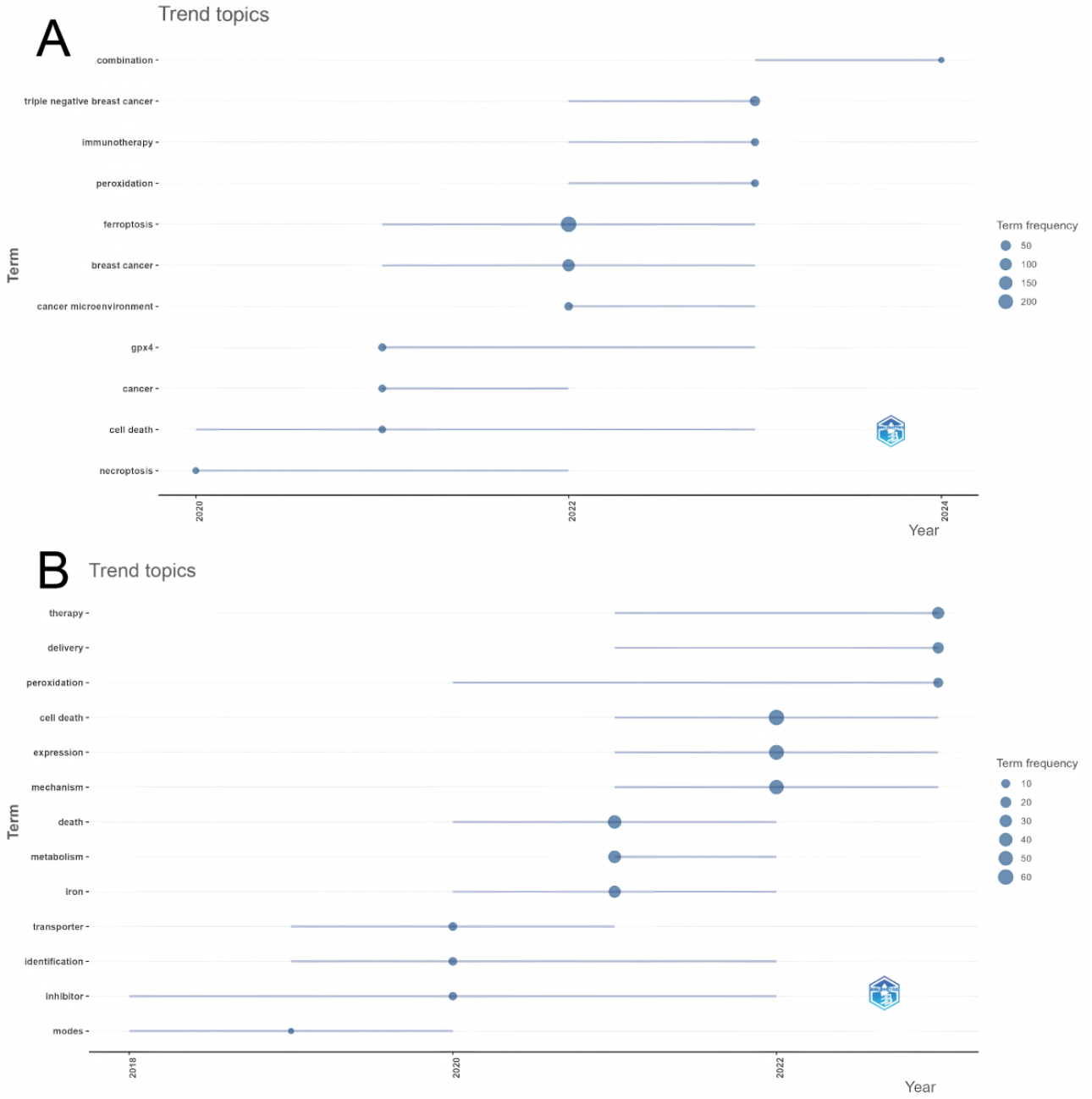
Cluster Analysis of Author Keywords and Analysis of Main Evolution Trends

After eliminating the core terms "breast cancer" and "ferroptosis," along with their synonyms, this study conducted a clustering analysis of the top 100 most frequent author keywords. The results indicate that these keywords can be categorized into four primary clusters (Multimedia Appendix 5). Specifically, Cluster 1 (in red) predominantly focuses on research topics such as "triple-negative breast cancer" (n=56), "prognosis model" (n=18), and "cancer microenvironment" (n=17). Cluster 2 (in green) centers its research around concepts such as "peroxidation" (n=13), "photothermal therapy" (n=12), and the general term "cancer" (n=11). Cluster 3 (in blue) revolves around high-frequency keywords, such as "reactive oxygen species" (n=17), "gpx4" (n=14), and "stress" (n=13). Finally, Cluster 4 (in orange) concentrates on areas related to "cell death" (n=10), "anticancer" (n=6), and "drug delivery" (n=5).

A deeper analysis was undertaken to examine the evolutionary patterns of high-frequency keywords over recent years, as presented in Multimedia Appendix 6A and Figure 7A. These trends reveal a progressive intensification of research focused on the concepts of cell death and glutathione peroxidase 4 (GPX4), indicating a growing interest and emphasis within the scientific community. It is noteworthy that from 2022 onwards, the frequency of keywords related to "cancer microenvironment," "triple-negative breast cancer," "immunotherapy," "combination therapy," and "peroxidation" increased significantly. Additionally, as witnessed from the time heatmap of the top 30 author keywords, keywords such as "erastin" and "autophagy" have maintained stable popularity in recent years (Multimedia Appendix 6; left panel).

XSL•FO
RenderX

Figure 7. Trend charts of keywords. (A) Trend chart of author keywords. (B) Trend chart of Keywords Plus.



Through a deep analysis of the thematic map data derived from the author keywords (Figure 8A), this study gained insights into the potential evolutionary directions in the research field of breast cancer and ferroptosis. The upper right quadrant of the map features motor themes, which incorporate the pivotal thematic terms including ferroptosis, breast cancer, and its subtype, triple-negative breast cancer (TNBC), underscoring their interconnectedness and centrality in the research landscape. The niche themes in the upper left corner encompass thematic keyword clusters such as anticancer and gene signatures. The emerging or declining themes in the lower left corner include topics such as chemodynamic therapy and photothermal therapy (PTT). Finally, in the lower right corner, the basic themes include thematic keywords such as cancer microenvironment, immunotherapy, and prognosis model.

XSL•FO
RenderX

Figure 8. Thematic maps of keywords. (A) Thematic map of author keywords. (B) Thematic map of Keywords Plus.



Cluster Analysis of Keywords Plus and Analysis of the Main Evolution Trends in Research

The statistical data revealed that some papers had missing keywords, with an omission rate of 19.64%. By increasing the likelihood of yielding more precise outcomes, keywords derived using the Keywords Plus feature can enhance the rigor and reliability of research findings. Therefore, the following section presents a detailed analysis of Keywords Plus.

During the clustering analysis of Keywords Plus, this study excluded the core terms “breast cancer” and “ferroptosis,” along with their synonyms, focusing instead on the analysis of the top 100 most frequent terms. The analysis revealed that these terms could be grouped into five clusters (Multimedia Appendix 7). Among these clusters, the keyword “cell death” stands out as a focal point, with a high frequency of 99 occurrences. Additionally, “expression” (n=53), “mechanism” (n=49),

XSL•FO
RenderX

"transporter" (n=19), and "stress" (n=9) were identified as core terms within their respective clusters.

This paper further analyses the evolution trends of keywords in the past and future. As shown in Multimedia Appendix 6 (right panel) and Figure 7B, the results indicate that researchers in the field of breast cancer and ferroptosis have consistently maintained a certain level of interest in the study of cell death in recent years. Additionally, newly emerged keywords such as "mechanism," "expression," and "nanoparticle" have been frequently observed in recent years.

Through the visualization presented in the thematic map (Figure 8B), the results indicate that the motor themes in the upper right quadrant consist of keywords such as "migration," "epithelial-mesenchymal transition (EMT)," and "negative breast cancer." The niche themes in the upper left quadrant encompass two core themes: "radiotherapy" and "progression." Within the radiotherapy category, keywords such as "alpha" emerge, while in the progression category, keywords such as "discovery" and "hallmarks" appear. The emerging or declining themes in the lower left quadrant reveal keywords such as "transporter," "flavonoids," and "modes." The basic themes in the lower right quadrant exhibit two main themes: "expression" and "cell death." Within the expression category, keywords such as "apoptosis" and "ferroptosis" are present, while in the cell death category, keywords such as "mechanism" and "nanoparticle" emerge.

Discussion

Principal Findings

Breast cancer, as a malignancy with a high mortality rate, still faces significant limitations in early diagnosis and late-stage treatment. In recent years, ferroptosis, a unique form of cell death triggered by cellular metabolic imbalance, has attracted widespread attention [11], underscoring the importance and potential of research in this area. This study systematically reviewed 387 papers related to breast cancer and ferroptosis, retrieved from the WOSCC database, published between 2016 and February 1, 2024. The analysis revealed that the number of publications in this field peaked in 2023, showing a significant growth trend. However, the annual citation frequency of related publications exhibited fluctuations, with the highest citation rate observed in 2017, followed by a gradual decline. Further analysis highlighted the prominent contribution of scholars such as Efferth T, with Chen Y having the highest citation frequency. In terms of institutions, Sun Yat-sen University led in the number of publications. Research output from China and the United States dominated this field, with increasing collaborative efforts between the two countries. While China had the highest total citation count, Germany recorded the highest average citation frequency per paper. The journal *Frontiers in Oncology* published the highest number of papers. Citation network analysis indicated that journals in "Molecular, Biology, Genetics" were central to the citation network. High-impact papers focused on the induction mechanisms of ferroptosis in breast cancer and the regulatory role of related proteins. Keyword analysis identified GPX4 as a central research trend. Additionally, breast cancer subtypes emerged as key drivers of ferroptosis research. Fundamental research themes pointed to the tumor immune microenvironment, prognostic models, and nanotherapy as future research priorities.

In-Depth Analysis of Publication Volume and Citation Counts in Breast Cancer and Ferroptosis Research

A comprehensive analysis of breast cancer and ferroptosis research reveals that publication volume and citation counts serve as crucial indicators for assessing the impact and development of this field. These metrics not only reflect the significance of research outcomes but also highlight emerging trends. An examination of frequently cited papers shows that key studies focus on the induction mechanisms of ferroptosis in breast cancer cells, protein regulation, and comparisons with other forms of cell death, aiming to discover novel therapeutic strategies. Specifically, the first study in this field published in 2016 demonstrated the effectiveness of siramesine and lapatinib in inducing ferroptosis in breast cancer cells [12], leading to high citation rates. Since then, exploration of the relationship between breast cancer and ferroptosis has deepened, with research becoming increasingly refined. TNBC, notable research by Yu et al [13] revealed that exosome-encapsulated erastin elevates ferritin levels, suggesting a new therapeutic approach. Hasegawa et al [14] further elucidated the intricate interplay between the cystine/glutamate antiporter, CD44v, and the MUC1-C oncoprotein in TNBC, enhancing our understanding of this subtype. Additionally, studies on the glutathione (GSH)-degrading enzyme CHAC1 have uncovered a novel mechanism of ferroptosis in TNBC via the GCN2-eIF2α axis [15].

The analysis of influential authors highlights those with high publication volumes and citation impacts. A distinguished team led by Efferth T has investigated the cytotoxicity of natural products, such as alkaloids and saponins, against multidrug-resistant cancers, focusing on their ability to induce various forms of cell death, including ferroptosis [16-21]. Their work provides important insights into overcoming drug resistance and developing novel anticancer therapies. High-impact authors, such as Chen Y and Gibson SB, have explored the therapeutic potential of ferroptosis in breast cancer, particularly through nanotechnology and drug design, contributing new perspectives to the field [22,23].

Institutional analysis reveals that Sun Yat-sen University and Zhejiang University lead in publication output in this area, reflecting their strong research foundations. Journals such as *Frontiers in Oncology*, *ACS Nano*, and *Cell Death & Disease* play pivotal roles in disseminating key findings and shaping the breast cancer and ferroptosis research landscape. By understanding the complex relationships between citing and cited journals, we uncover the distribution characteristics of core journals and knowledge flow paths within this interdisciplinary field, highlighting the contributions of journals from molecular biology, immunology, genetics, physics, materials science, and chemistry.

At the national level, China and the United States dominate in publications related to breast cancer and ferroptosis. China has made significant contributions to exploring the application of ferroptosis in cancer treatment, particularly through drugs, gene

XSL•FO
RenderX

regulation, and nanotechnology. The United States, on the other hand, focuses on the mechanisms and regulation of ferroptosis in cancer, including its impact on lipid composition, iron metabolism, and key genes influencing ferroptosis sensitivity [23-27]. Despite limited resources, research teams from resource-poor countries like India and Brazil have leveraged nanotechnology to propose novel therapeutic strategies [28,29]. While China ranks first in total citations, Germany and the United States have higher citation rates per paper, indicating greater research impact and quality, possibly due to higher research investment and international collaboration.

The publication trend shows robust development in this field, with 28 publications by February 1, 2024, suggesting that 2024 may surpass 2023 in research output. The slight decline in citation rates reflects the surge in publications and underscores the growing interest in exploring the interplay between breast cancer and ferroptosis.

This growing research body highlights ferroptosis in breast cancer as an emerging frontier in biomedical science. Leading scholars in the field continue to make significant contributions, and institutions and countries demonstrate high levels of cooperation and investment. These positive factors predict that the field of breast cancer and ferroptosis research will achieve major breakthroughs in the near future, opening up new avenues for the diagnosis and treatment of breast cancer with profound implications.

A Systematic Summary of Ferroptosis Mechanisms

Ferroptosis, characterized by its unique mechanism involving iron-dependent regulation through multiple pathways, plays a pivotal role in the onset and progression of various diseases. The core mechanism of ferroptosis lies in the imbalance between phospholipid peroxidation triggered by the oxidation of polyunsaturated fatty acids and the antioxidant system mediated by GPX4. When GPX4 activity decreases or GSH levels, as a cofactor, decline, phospholipid peroxidation products accumulate, leading to ferroptosis [5]. Methods to induce ferroptosis include the use of ferroptosis inducers (such as erastin and RSL3), inhibition of the system Xc-/GSH-GPX4 axis, and increased iron accumulation. Conversely, strategies to inhibit ferroptosis involve upregulating GPX4 expression, using ferroptosis inhibitors (such as ferrostatin-1), and enhancing the antioxidant system [7].

The signaling pathways and defense mechanisms of ferroptosis are complex and diverse. The initiation of iron toxicity involves multiple sources of reactive oxygen species (ROS), including iron-mediated Fenton reactions, mitochondrial ROS, and membrane-associated ROS driven by the NADPH oxidase (NOX) protein family. Cells defend against ferroptosis by upregulating antioxidant gene expression, repairing membrane damage, and clearing damaged organelles [30]. There is also a crosstalk between the mitochondrial dynamic regulatory network and ferroptosis, with mitochondrial dysfunction and damage potentially promoting ferroptosis [31]. By regulating lipid metabolism pathways, ferroptosis can affect cellular sensitivity to ferroptosis, thereby modulating its occurrence and development [32]. Furthermore, mechanisms such as epigenetic modifications and posttranslational modifications also play crucial roles in the regulation of ferroptosis [33].

Given that iron accumulation, oxidative stress, and lipid peroxidation are common features of various diseases, ferroptosis, as a unique cell death mechanism, may be a key link in the pathological processes of diverse diseases [34]. In neurodegenerative diseases, such as Parkinson disease and ischemic stroke, ferroptosis also plays a significant role [35]. In the field of cancer, ferroptosis is considered a potential antitumor strategy. For example, ferroptosis interacts with the tumor immune microenvironment, influencing tumor growth and metastasis [36]. Inhibiting ubiquitin-specific protease 8 can destabilize GPX4, thereby enhancing cancer cell sensitivity to ferroptosis, inhibiting tumor growth, and enhancing the efficacy of immunotherapy [37]. Additionally, ferroptosis is closely related to the death of plaque cells (such as vascular endothelial cells, macrophages, and vascular smooth muscle cells) and the development of atherosclerotic plaques in cardiovascular diseases [38]. Therefore, the application of ferroptosis in the treatment of various diseases, including cancer immunotherapy and neurodegenerative diseases, is gaining increasing attention.

Research on the regulatory mechanisms of ferroptosis in breast cancer indicates that its regulation involves multiple aspects, such as lipid metabolism and the immune microenvironment [39-42], and that ferroptosis plays a crucial role in the sensitivity and drug resistance of breast cancer to therapy [43-45]. With a deeper understanding of ferroptosis mechanisms, the methods for its detection are continuously evolving, including fluorescence probes, positron emission tomography imaging, and others, providing multiscale perspectives for ferroptosis research [46].

Future research should further elucidate the specific mechanisms of ferroptosis in the onset and progression of diseases, as well as its specific roles in the regulatory pathways, sensitivity, and resistance of breast cancer, providing new targets and strategies for breast cancer treatment.

Keyword Analysis: Key Hotspot GPX4 and Ferroptosis Regulation in Breast Cancer

An analysis of evolutionary trends has revealed the increasingly prominent importance of GPX4 in relevant academic research fields. Previous studies have illustrated that GPX4, as a core regulatory factor of ferroptosis, plays a crucial role in organisms by catalyzing the reduction of phospholipid peroxides to inhibit iron ion degeneration, thereby affecting the ferroptosis process [47,48]. It has been reported that RUNX1 intron transcript 1 (RUNX1-it1) is significantly overexpressed in breast cancer tissues, and it can block the impact of ferroptosis by increasing GPX4 expression, thereby promoting the occurrence of breast cancer [49].

On the other hand, Tubastatin A, as an efficient and selective Histone Deacetylase 6 inhibitor, has been confirmed by relevant studies as a novel ferroptosis inducer. It promotes ferroptosis by inhibiting the activity of GPX4, thereby enhancing the effect of cancer radiotherapy [50]. Chloramine T has also been found to inhibit the cell growth of breast cancer and induce ferroptosis

XSL•FO
RenderX

by targeting the GPX4 axis, indicating that it may become a potential new strategy for breast cancer treatment [51].

In addition, recent research has provided new insights into the regulatory mechanism of GPX4. As a crucial selenoprotein, the function of GPX4 in resisting ferroptosis depends greatly on the presence of its unique selenocysteine residue, which provides an important basis for understanding the mechanism of GPX4 [52]. The process of selenocysteine uptake mediated by low-density lipoprotein receptor-related protein 8 has been elucidated as a core mechanism in promoting GPX4 protein synthesis and effectively inhibiting ferroptosis [53,54].

Another study found that phospholipid-modifying enzymes MBOAT1 and MBOAT2, as ferroptosis inhibitors, regulate ferroptosis by remodeling the cellular phospholipid profile through the sex hormone signaling pathway [55]. These findings indicate that research on GPX4 and its related regulatory mechanisms is gradually deepening from phenotypic description to molecular mechanism analysis, offering unprecedented opportunities for the treatment of breast cancer.

Motor Themes Analysis in Author Keywords: Differential Sensitivity of Ferroptosis in Various Breast Cancer Subtypes Driving the Field's Development

Breast cancer is a highly heterogeneous disease, with different breast cancer cell types, characterized by the expression of estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and the proliferation marker Ki67, all exhibiting marked differences in sensitivity to ferroptosis. In the analysis of motor themes, findings related to ferroptosis sensitivity across breast cancer subtypes, including TNBC, have become well-established, closely interconnected with other emerging themes, and serve as a major driving force in the development of this field.

In ER-positive breast cancer, studies have shown that a combination of ferroptosis inducers, such as erastin and etoposide, can significantly enhance cellular sensitivity to ferroptosis by regulating IREB2/FPN1 expression and thus affecting iron metabolism, ultimately inducing ferroptosis [56]. Furthermore, ACSL4, a key regulator of ferroptosis, is upregulated in ER-positive breast cancer, promoting the accumulation of lipid peroxidation products and thereby enhancing cellular sensitivity to ferroptosis [57].

In contrast, sensitivity to ferroptosis in HER2-positive breast cancer may be regulated by multiple factors. For instance, inhibiting crVDAC3 can induce ferroptosis in breast cancer cells by reducing HSPB1 expression, thereby mediating resistance to deruxtecan in HER2-low breast cancer [58]. Additionally, m6A methylation modification of FGFR4 can inhibit ferroptosis, enhancing the resistance of HER2-positive breast cancer cells to anti-HER2 therapy [10]. Additionally, research has found that the inhibition of FGFR4 affects the expression of β-catenin/TCF4-SLC7A11 and FPN1, which has a significant impact on ferroptosis in HER2-positive breast cancer cells, thereby enhancing treatment sensitivity [10]. These studies suggest that the regulatory mechanisms of ferroptosis in HER2-positive breast cancer are more complex, involving the interaction of multiple signaling pathways.

In PR-positive breast cancer, although direct research on ferroptosis sensitivity is scarce, some indirect evidence suggests that PR may regulate the sensitivity of breast cancer cells to ferroptosis. Specifically, studies have reported that PR may indirectly regulate the sensitivity of TNBC to ferroptosis through metabolic pathways mediated by PR membrane component 1; however, these hypotheses await further experimental validation to establish their accuracy [59].

Relatively speaking, breast cancers that are negative for ER, PR, and HER2, especially TNBC, exhibit higher sensitivity to ferroptosis inducers. This characteristic may be closely related to the abnormal disruptions in iron and lipid metabolism in TNBC, which make TNBC cells more susceptible to ferroptosis inducers [60,61]. For example, the high expression level of ACSL4 in TNBC promotes the generation of lipid peroxides, a core step in the execution of ferroptosis [62]. Studies have also found that ferroptosis is enhanced through the combined action of GSH depletion and dihydroorotate dehydrogenase inhibitors [61]. Furthermore, the functions of key regulatory factors, such as the Xc-/GSH/GPX4 axis, ACSL4/LPCAT3 pathway, and nuclear factor erythroid 2-related factor 2 (NRF2) in ferroptosis, as well as their potential dysregulation mechanisms related to cancer cell survival and drug resistance, have been thoroughly explored [60]. Additionally, some natural compounds, such as artemisinin derivatives and traditional Chinese medicine extracts, have been found to inhibit the growth of TNBC cells by inducing ferroptosis [63].

Ki67, a commonly used proliferation marker, is often associated with higher proliferative activity and malignancy in breast cancer. Interestingly, breast cancer cells with high Ki67 expression also exhibit higher sensitivity to ferroptosis inducers [64]. However, research on the direct relationship between Ki67 expression and ferroptosis sensitivity is still limited. Future studies should further explore the role of Ki67 in the regulation of ferroptosis and whether it can serve as a biomarker for predicting the therapeutic effect of ferroptosis induction.

The motor themes analysis has provided a solid foundation for research in the area of breast cancer and ferroptosis. Continued investigation into the interactions between differential ferroptosis sensitivity across breast cancer subtypes holds the potential to uncover new therapeutic targets and biomarkers, ultimately driving the development of precision medicine and improving treatment outcomes and prognoses for patients with different breast cancer subtypes.

Analysis of Basic Themes in Breast Cancer and Ferroptosis Research: Immune Microenvironment and Prognostic Models Reveals Future Directions

The basic themes of immune microenvironment and prognostic models in breast cancer and ferroptosis research are closely interrelated and have led to several significant findings. For instance, in an insightful review by Shen et al [65], the authors explicitly point out the central role of iron in innate and adaptive immune responses, further emphasizing the potential positive influence of targeted regulation of iron metabolism on antitumor immunity and cancer treatment. Recent studies have demonstrated that T-cell–mediated immune responses [65] and

XSL•FO
RenderX

short-term acidosis-induced M1 macrophage polarization can effectively promote ferroptosis [66], thereby playing an active role in antitumor immunotherapy.

Research undertaken by Xu et al [67] has elegantly demonstrated that the development of ferroptosis-related gene (FRG) signatures holds significant promise as robust prognostic indicators for the immune microenvironment and therapeutic responsiveness. This approach shows great potential to inform and guide the design of future individualized treatment strategies, thereby enhancing precision medicine in breast cancer management.

A new lncRNA signal closely related to ferroptosis has been discovered, with its immune-related pathways significantly enriched in high-risk patients with breast cancer, demonstrating great potential for predicting the prognosis of patients with breast cancer [68]. Knocking down the hub gene SLC39A7, combined with TME scoring, can significantly affect the apoptosis and ferroptosis of cancer cells, with higher prognostic efficacy.

Meanwhile, epithelial cells and B cells exhibit higher ferroptosis scores, which are respectively related to the immune checkpoint blockade response and immune checkpoint blockade-independent gene expression, suggesting a connection between ferroptosis and the immune microenvironment [69]. It is noteworthy that research has accurately predicted the survival of patients with breast cancer through a novel prognostic model composed of nine FRGs [70].

Chen et al [71] used the Cancer Genome Atlas to identify 11 ferroptosis genes related to breast cancer prognosis, established a precise prognostic model, and analyzed therapeutic targets, thereby enhancing prognosis management. In TNBC, the subtype of breast cancer with the worst prognosis, research by Wu et al [72] evaluated the role of FRGs and found that they may affect prognostic models by regulating the tumor microenvironment, thus providing a solid theoretical basis for a clinically accurate prediction of TNBC prognosis. Furthermore, studies have evaluated the FRG AKR1C1, revealing its potential related to the immune-microenvironment, which may further affect the progress and prognosis of patients with breast cancer, making it a novel biomarker for the immune microenvironment and prognosis determination of breast cancer [73].

In conclusion, although significant progress has been made in the areas of immune microenvironment and prognostic models within the context of breast cancer and ferroptosis research, these fields remain underexplored and offer considerable potential for future investigation. These themes provide valuable directions for further research and offer essential guidance for researchers aiming to deepen their work in this area.

Innovative Applications and Strategies of Nanoparticles: Further Supplementing Future Research Directions in Breast Cancer and Ferroptosis

Through an in-depth analysis of the Keywords Plus basic themes, we further explore the future development directions of nanoparticles in the context of breast cancer and ferroptosis research. The application of nanoparticles in breast cancer treatment relies heavily on their unique physical and chemical properties to achieve the targeted delivery and controlled release of drugs, thereby enhancing therapeutic efficacy and reducing side effects [74]. Nanoparticles can exert their effects through various mechanisms that promote cell death, including ferroptosis [75]. Recent research has reported the use of a novel Fenton-independent pathway using photothermal nanozymes to overcome the limitations of the Fenton reaction. By using iron-containing hollow mesoporous Prussian blue (HMPB) nanocubes as an iron source to prepare HMPB@Lip, which combines PTT with nanozyme action, the induced ferroptosis through a non-Fenton reaction effectively ablates tumors, demonstrating the therapeutic potential of HMPB@Lip as a multifunctional nanozyme for ferroptosis. Furthermore, the development of Her2-DSG NPs nanostructures, which combine chemo-PTT with tumor microenvironment remodeling and immune activation, holds great promise for HER2-positive breast cancer [76].

Another study constructed nFeAPG nano-complex synergizing Fe^{3} with apigenin to enhance immune response through PTT, effectively controlling TNBC [71]. Additionally, the ICG@SANPs-cRGD nanomaterial-mediated combination of PTT/photodynamic therapy and immunotherapy has demonstrated its potential as a multifunctional platform for breast cancer treatment [77].

Furthermore, a search conducted on the Clinical Trials website revealed that there is only one clinical trial project related to breast cancer and ferroptosis, and that it is associated with nanocarriers (Multimedia Appendix 8). As of the search date, November 12, 2024, this project is still in the recruiting phase. Its focus is on the application of carbon nanoparticle-loaded iron (CNSI-Fe (II)) in the treatment of patients with advanced solid tumors, particularly those who have experienced treatment failure (including disease progression or intolerance) or lack of available standard therapeutic options (trial identifier: NCT06048367). The primary end point of this study is to assess the safety and tolerability of CNSI-Fe (II) throughout the study duration. The secondary end point, meanwhile, focuses on evaluating the pharmacokinetic profile of CNSI-Fe (II) in patients with advanced solid tumors.

Nanotechnology, as a fundamental topic in the research fields of breast cancer and ferroptosis, is still in its relatively early stages of development. Nonetheless, these research achievements have not only significantly expanded the application scope of nanotechnology in breast cancer treatment but have also highlighted its close connection with breast cancer therapy and identified numerous promising therapeutic research directions that await further exploration. More importantly, they have pointed out new avenues for future research endeavors and provided invaluable guidance for investigative pursuits.

Limitations

This study represents the first comprehensive bibliometric analysis of 387 academic papers related to breast cancer and ferroptosis sourced from the WOSCC database. By applying multiple analytical approaches, it thoroughly investigates the research hotspots and future trends in this field, offering a

XSL•FO
RenderX

comprehensive and systematic overview of the latest developments in this area. These strengths significantly enhance the academic value and practical relevance of the study, providing essential guidance for future advancements in the field of breast cancer and ferroptosis.

However, several limitations of our methodology must be acknowledged. First, although the WOSCC database is vast, it was not specifically designed for bibliometric analysis and relies on a single data source, which may not capture all relevant literature present in other databases (such as PubMed, Scopus, and Google Scholar), potentially introducing bias. Future research should integrate multiple databases to mitigate this issue. Second, our focus on English-language publications limits the comprehensiveness of the study by excluding high-quality non-English research. To address this, future studies should adopt multilingual search strategies. Third, the exclusion of nonpaper publications (eg, reviews and conference papers) constrains the scope of research outputs considered. To obtain a more holistic view, future work should adopt more inclusive selection criteria. Finally, given the rapid updates in databases, there is a risk of overlooking recent advances in the field's literature. Periodic updates to bibliometric analyses are recommended to ensure timely and accurate capture of the latest trends and developments in breast cancer and ferroptosis research.

Conclusions

Ferroptosis holds significant value and potential in breast cancer research. Through bibliometric analysis, this study suggests that enhancing collaboration and communication among authors, institutions, and countries, while leveraging the interdisciplinary nature of journal citations, can facilitate advances in fundamental research on ferroptosis in breast cancer. Specifically, this paper reviews the progress in understanding the mechanisms of ferroptosis, providing guidance for breast cancer research. It emphasizes the sensitivity of different breast cancer subtypes to ferroptosis as a driving force in the current research landscape and highlights the pivotal role of GPX4 in the regulatory mechanisms of ferroptosis in breast cancer as a key focus. Additionally, the study identifies the importance of the tumor microenvironment, immunotherapy, and prognostic models for future development. Although the application of nanoparticles in breast cancer treatment is in its preliminary stages, it has already shown promising prospects and serves as a rich supplement to the future directions. By elucidating the current status and future trends of ferroptosis in breast cancer research, this study provides direction for future research endeavors and promotes innovation in breast cancer treatment strategies.

Acknowledgments

All authors would like to thank the "Guangxi Zhuang Autonomous Region Clinical Medicine Research Center for Molecular Pathology and Intelligent Pathology Precision Diagnosis" for providing relevant technical support. This study was supported by the Guangxi Medical University Teacher Teaching Ability Development Project (2202JFA02), and the Guangxi Medical University Undergraduate Education and Teaching Reform Project (2023Y05).

Data Availability

The datasets generated during or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

JYL, YLD, SYL, SYC, RQH, DYQ, BTC, GC, XY, and WP contributed to the study conception and design. Material preparation, data collection, and analysis were performed by JYL, YLD, SYL, SYC, DYQ, and BTC. The first draft of the manuscript was written by JYL and YLD. SYL, RQH, GC, XY, and WP revised the paper. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key information of the included studies.
[PNG File , 377 KB - ijmr_v14i1e66286_app1.png]

Multimedia Appendix 2
Top 10 most frequently cited articles in research on ferroptosis and breast cancer.
[XLSX File (Microsoft Excel File), 11 KB - ijmr_v14i1e66286_app2.xlsx]

Multimedia Appendix 3
Top 10 authors with the most citations in ferroptosis and breast cancer research within the field.
[XLSX File (Microsoft Excel File), 276 KB - ijmr_v14i1e66286_app3.xlsx]

XSL•FO
RenderX

Multimedia Appendix 4
Top 10 countries ranked by total citations and average article citations in the field of ferroptosis and breast cancer research.
[XLSX File (Microsoft Excel File), 10 KB - ijmr_v14i1e66286_app4.xlsx]

Multimedia Appendix 5
Author keywords clustering visualisation in the field of breast cancer and ferroptosis research.
[PNG File , 915 KB - ijmr_v14i1e66286_app5.png]

Multimedia Appendix 6
Heatmap of keyword analysis.
[PNG File , 742 KB - ijmr_v14i1e66286_app6.png]

Multimedia Appendix 7
Visualisation of keywords plus clustering in breast cancer and ferroptosis research fields.
[PNG File , 880 KB - ijmr_v14i1e66286_app7.png]

Multimedia Appendix 8
Summary of clinical trials in the field of ferroptosis and breast cancer research (n=1).
[XLSX File (Microsoft Excel File), 13 KB - ijmr_v14i1e66286_app8.xlsx]

References

1. Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024;74(3):229-263 [FREE Full text] [doi: 10.3322/caac.21834] [Medline: 38572751]
2. Arnold M, Morgan E, Rumgay H, Mafra A, Singh D, Laversanne M, et al. Current and future burden of breast cancer: global statistics for 2020 and 2040. Breast 2022;66:15-23 [FREE Full text] [doi: 10.1016/j.breast.2022.08.010] [Medline: 36084384]
3. Admoun C, Mayrovitz HN. The eiology of breast cancer. In: Mayrovitz HN, editor. Breast Cancer [Internet]. Brisbane (AU): Exon Publications; 2022.
4. Breast Cancer Expert Committee of National Cancer Quality Control Center, Breast Cancer Expert Committee of China Anti-Cancer Association, Cancer Drug Clinical Research Committee of China Anti-Cancer Association. Guidelines for clinical diagnosis and treatment of advanced breast cancer in China (2022 edition). Zhonghua Zhong Liu Za Zhi 2022;44(12):1262-1287. [doi: 10.3760/cma.j.cn112152-20221007-00680] [Medline: 36575780]
5. Jiang X, Stockwell BR, Conrad M. Ferroptosis: mechanisms, biology and role in disease. Nat Rev Mol Cell Biol 2021;22(4):266-282 [FREE Full text] [doi: 10.1038/s41580-020-00324-8] [Medline: 33495651]
6. Li J, Cao F, Yin HL, Huang ZJ, Lin ZT, Mao N, et al. Ferroptosis: past, present and future. Cell Death Dis 2020;11(2):88 [FREE Full text] [doi: 10.1038/s41419-020-2298-2] [Medline: 32015325]
7. Tang D, Chen X, Kang R, Kroemer G. Ferroptosis: molecular mechanisms and health implications. Cell Res 2021;31(2):107-125 [FREE Full text] [doi: 10.1038/s41422-020-00441-1] [Medline: 33268902]
8. Dong X, Li Y, Sheng X, Zhou W, Sun A, Dai H. Mitochondria-related signaling pathways involved in breast cancer regulate ferroptosis. Genes Dis 2024;11(1):358-366 [FREE Full text] [doi: 10.1016/j.gendis.2023.03.019] [Medline: 37588231]
9. Xu Z, Wang X, Sun W, Xu F, Kou H, Hu W, et al. RelB-activated GPX4 inhibits ferroptosis and confers tamoxifen resistance in breast cancer. Redox Biol 2023;68:102952 [FREE Full text] [doi: 10.1016/j.redox.2023.102952] [Medline: 37944384]
10. Zou Y, Zheng S, Xie X, Ye F, Hu X, Tian Z, et al. N6-methyladenosine regulated FGFR4 attenuates ferroptotic cell death in recalcitrant HER2-positive breast cancer. Nat Commun 2022;13(1):2672 [FREE Full text] [doi: 10.1038/s41467-022-30217-7] [Medline: 35562334]
11. Mao C, Wang M, Zhuang L, Gan B. Metabolic cell death in cancer: ferroptosis, cuproptosis, disulfidptosis, and beyond. Protein Cell 2024;15(9):642-660 [FREE Full text] [doi: 10.1093/procel/pwae003] [Medline: 38428031]
12. Ma S, Henson ES, Chen Y, Gibson SB. Ferroptosis is induced following siramesine and lapatinib treatment of breast cancer cells. Cell Death Dis 2016;7(7):e2307 [FREE Full text] [doi: 10.1038/cddis.2016.208] [Medline: 27441659]
13. Yu M, Gai C, Li Z, Ding D, Zheng J, Zhang W, et al. Targeted exosome-encapsulated erastin induced ferroptosis in triple negative breast cancer cells. Cancer Sci 2019;110(10):3173-3182 [FREE Full text] [doi: 10.1111/cas.14181] [Medline: 31464035]
14. Hasegawa M, Takahashi H, Rajabi H, Alam M, Suzuki Y, Yin L, et al. Functional interactions of the cystine/glutamate antiporter, CD44v and MUC1-C oncoprotein in triple-negative breast cancer cells. Oncotarget 2016;7(11):11756-11769 [FREE Full text] [doi: 10.18632/oncotarget.7598] [Medline: 26930718]

XSL•FO
RenderX

15. Chen MS, Wang SF, Hsu CY, Yin PH, Yeh TS, Lee HC, et al. CHAC1 degradation of glutathione enhances cystine-starvation-induced necroptosis and ferroptosis in human triple negative breast cancer cells via the GCN2-eIF2α-ATF4 pathway. Oncotarget 2017;8(70):114588-114602 [FREE Full text] [doi: 10.18632/oncotarget.23055] [Medline: 29383104]
16. Mbaveng AT, Noulala CGT, Samba ARM, Tankeo SB, Abdelfatah S, Fotso GW, et al. The alkaloid, soyauxinium chloride, displays remarkable cytotoxic effects towards a panel of cancer cells, inducing apoptosis, ferroptosis and necroptosis. Chem Biol Interact 2021;333:109334. [doi: 10.1016/j.cbi.2020.109334] [Medline: 33245930]
17. Mbaveng AT, Bitchagno GTM, Kuete V, Tane P, Efferth T. Cytotoxicity of ungeremine towards multi-factorial drug resistant cancer cells and induction of apoptosis, ferroptosis, necroptosis and autophagy. Phytomedicine 2019;60:152832. [doi: 10.1016/j.phymed.2019.152832] [Medline: 31031043]
18. Mbaveng AT, Fotso GW, Ngnintedo D, Kuete V, Ngadjui BT, Keumedjio F, et al. Cytotoxicity of epunctanone and four other phytochemicals isolated from the medicinal plants Garcinia epunctata and Ptycholobium contortum towards multi-factorial drug resistant cancer cells. Phytomedicine 2018;48:112-119. [doi: 10.1016/j.phymed.2017.12.016] [Medline: 30195869]
19. Mbaveng AT, Ndontsa BL, Kuete V, Nguekeu YMM, Çelik İ, Mbouangouere R, et al. A naturally occuring triterpene saponin ardisiacrispin B displayed cytotoxic effects in multi-factorial drug resistant cancer cells via ferroptotic and apoptotic cell death. Phytomedicine 2018;43:78-85. [doi: 10.1016/j.phymed.2018.03.035] [Medline: 29747757]
20. Mbaveng AT, Chi GF, Bonsou IN, Abdelfatah S, Tamfu AN, Yeboah EMO, et al. N-acetylglycoside of oleanolic acid (aridanin) displays promising cytotoxicity towards human and animal cancer cells, inducing apoptotic, ferroptotic and necroptotic cell death. Phytomedicine 2020;76:153261. [doi: 10.1016/j.phymed.2020.153261] [Medline: 32559584]
21. Mbaveng AT, Damen F, Guefack MF, Tankeo SB, Abdelfatah S, Bitchagno GTM, et al. 8,8-bis-(Dihydroconiferyl)-diferulate displayed impressive cytotoxicity towards a panel of human and animal cancer cells. Phytomedicine 2020;70:153215. [doi: 10.1016/j.phymed.2020.153215] [Medline: 32388040]
22. Ma S, Dielschneider RF, Henson ES, Xiao W, Choquette TR, Blankstein AR, et al. Ferroptosis and autophagy induced cell death occur independently after siramesine and lapatinib treatment in breast cancer cells. PLoS One 2017;12(8):e0182921 [FREE Full text] [doi: 10.1371/journal.pone.0182921] [Medline: 28827805]
23. Chen Y, Yao Z, Liu P, Hu Q, Huang Y, Ping L, et al. A self-assembly nano-prodrug for triple-negative breast cancer combined treatment by ferroptosis therapy and chemotherapy. Acta Biomater 2023;159:275-288. [doi: 10.1016/j.actbio.2023.01.050] [Medline: 36709836]
24. Yi J, Zhu J, Wu J, Thompson CB, Jiang X. Oncogenic activation of PI3K-AKT-mTOR signaling suppresses ferroptosis via SREBP-mediated lipogenesis. Proc Natl Acad Sci U S A 2020;117(49):31189-31197 [FREE Full text] [doi: 10.1073/pnas.2017152117] [Medline: 33229547]
25. Brown CW, Amante JJ, Chhoy P, Elaimy AL, Liu H, Zhu LJ, et al. Prominin2 drives ferroptosis resistance by stimulating iron export. Dev Cell 2019;51(5):575-586.e4 [FREE Full text] [doi: 10.1016/j.devcel.2019.10.007] [Medline: 31735663]
26. Liu T, Jiang L, Tavana O, Gu W. The deubiquitylase OTUB1 mediates ferroptosis via stabilization of SLC7A11. Cancer Res 2019;79(8):1913-1924 [FREE Full text] [doi: 10.1158/0008-5472.CAN-18-3037] [Medline: 30709928]
27. Llabani E, Hicklin RW, Lee HY, Motika SE, Crawford LA, Weerapana E, et al. Diverse compounds from pleuromutilin lead to a thioredoxin inhibitor and inducer of ferroptosis. Nat Chem 2019;11(6):521-532 [FREE Full text] [doi: 10.1038/s41557-019-0261-6] [Medline: 31086302]
28. Szwed M, Torgersen ML, Kumari RV, Yadava SK, Pust S, Iversen TG, et al. Biological response and cytotoxicity induced by lipid nanocapsules. J Nanobiotechnol 2020;18(1):5 [FREE Full text] [doi: 10.1186/s12951-019-0567-y] [Medline: 31907052]
29. Zhu J, Dai P, Liu F, Li Y, Qin Y, Yang Q, et al. Upconverting nanocarriers enable triggered microtubule inhibition and concurrent ferroptosis induction for selective treatment of triple-negative breast cancer. Nano Lett 2020;20(9):6235-6245. [doi: 10.1021/acs.nanolett.0c00502] [Medline: 32804509]
30. Liu J, Kang R, Tang D. Signaling pathways and defense mechanisms of ferroptosis. FEBS J 2022;289(22):7038-7050 [FREE Full text] [doi: 10.1111/febs.16059] [Medline: 34092035]
31. Li J, Jia YC, Ding YX, Bai J, Cao F, Li F. The crosstalk between ferroptosis and mitochondrial dynamic regulatory networks. Int J Biol Sci 2023;19(9):2756-2771 [FREE Full text] [doi: 10.7150/ijbs.83348] [Medline: 37324946]
32. Pope LE, Dixon SJ. Regulation of ferroptosis by lipid metabolism. Trends Cell Biol 2023;33(12):1077-1087. [doi: 10.1016/j.tcb.2023.05.003] [Medline: 37407304]
33. Wang Y, Hu J, Wu S, Fleishman JS, Li Y, Xu Y, et al. Targeting epigenetic and posttranslational modifications regulating ferroptosis for the treatment of diseases. Signal Transduction Targeted Ther 2023;8(1):449 [FREE Full text] [doi: 10.1038/s41392-023-01720-0] [Medline: 38072908]
34. Mahoney-Sánchez L, Bouchaoui H, Ayton S, Devos D, Duce JA, Devedjian JC. Ferroptosis and its potential role in the physiopathology of Parkinson's disease. Prog Neurobiol 2021;196:101890 [FREE Full text] [doi: 10.1016/j.pneurobio.2020.101890] [Medline: 32726602]
35. Hu X, Bao Y, Li M, Zhang W, Chen C. The role of ferroptosis and its mechanism in ischemic stroke. Exp Neurol 2024;372:114630. [doi: 10.1016/j.expneurol.2023.114630] [Medline: 38056585]

XSL•FO
RenderX

36. Chen X, Kang R, Kroemer G, Tang D. Broadening horizons: the role of ferroptosis in cancer. Nat Rev Clin Oncol 2021;18(5):280-296. [doi: 10.1038/s41571-020-00462-0] [Medline: 33514910]
37. Li H, Sun Y, Yao Y, Ke S, Zhang N, Xiong W, et al. USP8-governed GPX4 homeostasis orchestrates ferroptosis and cancer immunotherapy. Proc Natl Acad Sci U S A 2024;121(16):e2315541121. [doi: 10.1073/pnas.2315541121] [Medline: 38598341]
38. Xu X, Xu XD, Ma MQ, Liang Y, Cai YB, Zhu ZX, et al. The mechanisms of ferroptosis and its role in atherosclerosis. Biomed Pharmacother 2024;171:116112 [FREE Full text] [doi: 10.1016/j.biopha.2023.116112] [Medline: 38171246]
39. Ding Y, Chen X, Liu C, Ge W, Wang Q, Hao X, et al. Identification of a small molecule as inducer of ferroptosis and apoptosis through ubiquitination of GPX4 in triple negative breast cancer cells. J Hematol Oncol 2021;14(1):19 [FREE Full text] [doi: 10.1186/s13045-020-01016-8] [Medline: 33472669]
40. Wang Z, Li R, Hou N, Zhang J, Wang T, Fan P, et al. PRMT5 reduces immunotherapy efficacy in triple-negative breast cancer by methylating KEAP1 and inhibiting ferroptosis. J Immunother Cancer 2023;11(6):e006890 [FREE Full text] [doi: 10.1136/jitc-2023-006890] [Medline: 37380368]
41. Doll S, Proneth B, Tyurina YY, Panzilius E, Kobayashi S, Ingold I, et al. ACSL4 dictates ferroptosis sensitivity by shaping cellular lipid composition. Nat Chem Biol 2017;13(1):91-98 [FREE Full text] [doi: 10.1038/nchembio.2239] [Medline: 27842070]
42. Zhao Y, Liu Z, Liu G, Zhang Y, Liu S, Gan D, et al. Neutrophils resist ferroptosis and promote breast cancer metastasis through aconitate decarboxylase 1. Cell Metab 2023;35(10):1688-1703.e10 [FREE Full text] [doi: 10.1016/j.cmet.2023.09.004] [Medline: 37793345]
43. Shen M, Cao S, Long X, Xiao L, Yang L, Zhang P, et al. DNAJC12 causes breast cancer chemotherapy resistance by repressing doxorubicin-induced ferroptosis and apoptosis via activation of AKT. Redox Biol 2024;70:103035 [FREE Full text] [doi: 10.1016/j.redox.2024.103035] [Medline: 38306757]
44. Yang J, Zhou Y, Xie S, Wang J, Li Z, Chen L, et al. Metformin induces ferroptosis by inhibiting UFMylation of SLC7A11 in breast cancer. J Exp Clin Cancer Res 2021;40(1):206 [FREE Full text] [doi: 10.1186/s13046-021-02012-7] [Medline: 34162423]
45. Zhu Z, Shen H, Xu J, Fang Z, Wo G, Ma Y, et al. GATA3 mediates doxorubicin resistance by inhibiting CYB5R2-catalyzed iron reduction in breast cancer cells. Drug Resist Updates 2023;69:100974 [FREE Full text] [doi: 10.1016/j.drup.2023.100974] [Medline: 37230023]
46. Zeng F, Nijiati S, Tang L, Ye J, Zhou Z, Chen X. Ferroptosis detection: from approaches to applications. Angew Chem Int Ed Engl 2023;62(35):e202300379. [doi: 10.1002/anie.202300379] [Medline: 36828775]
47. Liang D, Feng Y, Zandkarimi F, Wang H, Zhang Z, Kim J, et al. Ferroptosis surveillance independent of GPX4 and differentially regulated by sex hormones. Cell 2023;186(13):2748-2764.e22 [FREE Full text] [doi: 10.1016/j.cell.2023.05.003] [Medline: 37267948]
48. Luo M, Yan J, Hu X, Li H, Li H, Liu Q, et al. Targeting lipid metabolism for ferroptotic cancer therapy. Apoptosis 2023;28(1-2):81-107. [doi: 10.1007/s10495-022-01795-0] [Medline: 36399287]
49. Wang S, Wang Y, Li Q, Zeng K, Li X, Feng X. RUNX1-IT1 favors breast cancer carcinogenesis through regulation of IGF2BP1/GPX4 axis. Discov Oncol 2023;14(1):42 [FREE Full text] [doi: 10.1007/s12672-023-00652-z] [Medline: 37036576]
50. Liu S, Zhang HL, Li J, Ye ZP, Du T, Li LC, et al. Tubastatin a potently inhibits GPX4 activity to potentiate cancer radiotherapy through boosting ferroptosis. Redox Biol 2023;62:102677 [FREE Full text] [doi: 10.1016/j.redox.2023.102677] [Medline: 36989572]
51. Li H, Liu W, Zhang X, Wu F, Sun D, Wang Z. Ketamine suppresses proliferation and induces ferroptosis and apoptosis of breast cancer cells by targeting KAT5/GPX4 axis. Biochem Biophys Res Commun 2021;585:111-116. [doi: 10.1016/j.bbrc.2021.11.029] [Medline: 34800882]
52. Ingold I, Berndt C, Schmitt S, Doll S, Poschmann G, Buday K, et al. Selenium utilization by GPX4 is required to prevent hydroperoxide-induced ferroptosis. Cell 2018;172(3):409-422.e21 [FREE Full text] [doi: 10.1016/j.cell.2017.11.048] [Medline: 29290465]
53. Alborzinia H, Chen Z, Yildiz U, Freitas FP, Vogel FCE, Varga JP, et al. LRP8-mediated selenocysteine uptake is a targetable vulnerability in MYCN-amplified neuroblastoma. EMBO Mol Med 2023;15(8):e18014 [FREE Full text] [doi: 10.15252/emmm.202318014] [Medline: 37435859]
54. Li Z, Ferguson L, Deol KK, Roberts MA, Magtanong L, Hendricks JM, et al. Ribosome stalling during selenoprotein translation exposes a ferroptosis vulnerability. Nat Chem Biol 2022;18(7):751-761 [FREE Full text] [doi: 10.1038/s41589-022-01033-3] [Medline: 35637349]
55. NA. Sex hormone signaling suppresses ferroptosis via phospholipid remodeling. Cancer Discovery 2023;13(8):1759. [doi: 10.1158/2159-8290.CD-RW2023-091] [Medline: 37326388]
56. Ozkan E, Bakar-Ates F. Etoposide in combination with erastin synergistically altered iron homeostasis and induced ferroptotic cell death through regulating IREB2/FPN1 expression in estrogen receptor positive-breast cancer cells. Life Sci 2023;312:121222. [doi: 10.1016/j.lfs.2022.121222] [Medline: 36442526]

XSL•FO
RenderX

57. Hou J, Jiang C, Wen X, Li C, Xiong S, Yue T, et al. ACSL4 as a potential target and biomarker for anticancer: from molecular mechanisms to clinical therapeutics. Front Pharmacol 2022;13:949863 [FREE Full text] [doi: 10.3389/fphar.2022.949863] [Medline: 35910359]
58. Zou Y, Yang A, Chen B, Deng X, Xie J, Dai D, et al. crVDAC3 alleviates ferroptosis by impeding HSPB1 ubiquitination and confers trastuzumab deruxtecan resistance in HER2-low breast cancer. Drug Resist Updates 2024;77:101126. [doi: 10.1016/j.drup.2024.101126] [Medline: 39243601]
59. Zhao Y, Ruan X, Cheng J, Xu X, Gu M, Mueck AO. PGRMC1 promotes triple-negative breast cancer cell growth via suppressing ferroptosis. Climacteric 2023;26(2):135-142. [doi: 10.1080/13697137.2023.2170225] [Medline: 36724820]
60. Mokhtarpour K, Razi S, Rezaei N. Ferroptosis as a promising targeted therapy for triple negative breast cancer. Breast Cancer Res Treat 2024;207(3):497-513. [doi: 10.1007/s10549-024-07387-7] [Medline: 38874688]
61. Zhou TJ, Zhang MM, Liu DM, Huang LL, Yu HQ, Wang Y, et al. Glutathione depletion and dihydroorotate dehydrogenase inhibition actuated ferroptosis-augment to surmount triple-negative breast cancer. Biomaterials 2024;305:122447. [doi: 10.1016/j.biomaterials.2023.122447] [Medline: 38154441]
62. Monaco ME. ACSL4: biomarker, mediator and target in quadruple negative breast cancer. Oncotarget 2023;14:563-575 [FREE Full text] [doi: 10.18632/oncotarget.28453] [Medline: 37306503]
63. Wang Y, Yuan X, Ren M, Wang Z. Ferroptosis: a new research direction of artemisinin and its derivatives in anti-cancer treatment. Am J Chin Med 2024;52(1):161-181. [doi: 10.1142/S0192415X24500071] [Medline: 38328829]
64. Yeh NT, Lin TC, Liu IJ, Hu SH, Hsu TC, Chin HY, et al. Hyaluronic acid-covered ferric ion-rich nanobullets with high zoledronic acid payload for breast tumor-targeted chemo/chemodynamic therapy. Int J Biol Macromol 2024;279(Pt 3):135271. [doi: 10.1016/j.ijbiomac.2024.135271] [Medline: 39233170]
65. Shen L, Zhou Y, He H, Chen W, Lenahan C, Li X, et al. Crosstalk between macrophages, T cells, and iron metabolism in tumor microenvironment. Oxid Med Cell Longevity 2021;2021:8865791 [FREE Full text] [doi: 10.1155/2021/8865791] [Medline: 33628389]
66. Xiong H, Zhai Y, Meng Y, Wu Z, Qiu A, Cai Y, et al. Acidosis activates breast cancer ferroptosis through ZFAND5/SLC3A2 signaling axis and elicits M1 macrophage polarization. Cancer Lett 2024;587:216732. [doi: 10.1016/j.canlet.2024.216732] [Medline: 38360142]
67. Xu Y, Du Y, Zheng Q, Zhou T, Ye B, Wu Y, et al. Identification of ferroptosis-related prognostic signature and subtypes related to the immune microenvironment for breast cancer patients receiving neoadjuvant chemotherapy. Front Immunol 2022;13:895110 [FREE Full text] [doi: 10.3389/fimmu.2022.895110] [Medline: 35603151]
68. Zhang K, Ping L, Du T, Liang G, Huang Y, Li Z, et al. A ferroptosis-related lncRNAs signature predicts prognosis and immune microenvironment for breast cancer. Front Mol Biosci 2021;8:678877 [FREE Full text] [doi: 10.3389/fmolb.2021.678877] [Medline: 34164433]
69. Qiu Z, Deng C, Zhou F, Chen Y, Chen X, Liu X, et al. Ferroptosis heterogeneity within the tumor microenvironment revealed a genetic blueprint of breast cancer. Environ Toxicol 2024;39(5):2741-2752. [doi: 10.1002/tox.24142] [Medline: 38251953]
70. Lu YJ, Gong YJ, Li WY, Zhao C, Guo F. The prognostic significance of a novel ferroptosis-related gene model in breast cancer. Ann Transl Med 2022;10(4):184 [FREE Full text] [doi: 10.21037/atm-22-479] [Medline: 35280394]
71. Chen R, Jiang Z, Cheng Y, Ye J, Li S, Xu Y, et al. Multifunctional iron-apigenin nanocomplex conducting photothermal therapy and triggering augmented immune response for triple negative breast cancer. Int J Pharm 2024;655:124016. [doi: 10.1016/j.ijpharm.2024.124016] [Medline: 38503397]
72. Wu S, Pan R, Lu J, Wu X, Xie J, Tang H, et al. Development and verification of a prognostic ferroptosis-related gene model in triple-negative breast cancer. Front Oncol 2022;12:896927 [FREE Full text] [doi: 10.3389/fonc.2022.896927] [Medline: 35719954]
73. Zhang Z, Qiu X, Yan Y, Liang Q, Cai Y, Peng B, et al. Evaluation of ferroptosis-related gene as a novel biomarker associated with the immune microenvironment and prognosis in breast cancer. Int J Gen Med 2021;14:6189-6200 [FREE Full text] [doi: 10.2147/IJGM.S329031] [Medline: 34611432]
74. Marshall SK, Angsantikul P, Pang Z, Nasongkla N, Hussen RSD, Thamphiwatana SD. Biomimetic targeted theranostic nanoparticles for breast cancer treatment. Molecules 2022;27(19):6473 [FREE Full text] [doi: 10.3390/molecules27196473] [Medline: 36235009]
75. Herdiana Y, Sriwidodo S, Sofian FF, Wilar G, Diantini A. Nanoparticle-based antioxidants in stress signaling and programmed cell death in breast cancer treatment. Molecules 2023;28(14):5305 [FREE Full text] [doi: 10.3390/molecules28145305] [Medline: 37513179]
76. Zheng D, Wan C, Yang H, Xu L, Dong Q, Du C, et al. Her2-targeted multifunctional nano-theranostic platform mediates tumor microenvironment remodeling and immune activation for breast cancer treatment. Int J Nanomed 2020;15:10007-10028 [FREE Full text] [doi: 10.2147/IJN.S271213] [Medline: 33376321]
77. Kong C, Xu B, Qiu G, Wei M, Zhang M, Bao S, et al. Multifunctional nanoparticles-mediated PTT/PDT synergistic immune activation and antitumor activity combined with anti-PD-L1 immunotherapy for breast cancer treatment. Int J Nanomed 2022;17:5391-5411 [FREE Full text] [doi: 10.2147/IJN.S373282] [Medline: 36419717]

XSL•FO
RenderX

Abbreviations

ER: estrogen receptor
FRG: ferroptosis-related gene
GPX4: glutathione peroxidase 4
GSH: glutathione
HER2: human epidermal growth factor receptor 2
HMPB: hollow mesoporous Prussian blue
PPT: photothermal therapy
PR: progesterone receptor
ROS: reactive oxygen species
TNBC: triple-negative breast cancer
WOSCC: Web of Science Core Collection

Edited by T de Azevedo Cardoso; submitted 09.09.24; peer-reviewed by Y Zou, C Mao, Z Li; comments to author 01.10.24; revised version received 26.11.24; accepted 05.12.24; published 26.02.25.

Please cite as:
Luo JY, Deng YL, Lu SY, Chen SY, He RQ, Qin DY, Chi BT, Chen G, Yang X, Peng W
Current Status and Future Directions of Ferroptosis Research in Breast Cancer: Bibliometric Analysis
Interact J Med Res 2025;14:e66286
URL: https://www.i-jmr.org/2025/1/e66286
doi:10.2196/66286
PMID:

©Jia-Yuan Luo, Yu-Long Deng, Shang-Yi Lu, Si-Yan Chen, Rong-Quan He, Di-Yuan Qin, Bang-Teng Chi, Gang Chen, Xia Yang, Wei Peng. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.02.2025. 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

Process Evaluations of Interventions for the Prevention of Type 2 Diabetes in Women With Gestational Diabetes Mellitus: Systematic Review

Iklil Iman Mohd Sa'id[1*], PhD; Natasha Hotung[2*], BA(Hons), MA, MSc; Madeleine Benton[2], CPychol, PhD; Iliatha Papachristou Nadal[2], CPsychol, PhD; Anisah Baharom[1], PhD; Matthew Prina[3], PhD; Barakatun Nisak Mohd Yusof[1], PhD; Kimberley Goldsmith[2], PhD; Samantha Birts[4], BSc; Ching Siew Mooi[1], MD, MMed Fam, PhD; Angus Forbes[2], BSc, MSc, PhD; Khalida Ismail[2], MBBS, PhD; Boon How Chew[1], MD, PhD

[1]Department of Family Medicine, Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Selangor, Malaysia

[2]Department of Psychological Medicine, King's College London, London, United Kingdom

[3]Newcastle University, Newcastle upon Tyne, United Kingdom

[4]Department of Psychological Medicine, King's College London, Leeds, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Khalida Ismail, MBBS, PhD
Department of Psychological Medicine
King's College London
Institute of Psychiatry Psychology and Neuroscience
Denmark Hill Campus, King’s College London, 16 De Crespigny Park
London, SE5 8AB
United Kingdom
Phone: 44 7905985068
Email: khalida.2.ismail@kcl.ac.uk

Abstract

Background: Gestational diabetes mellitus (GDM) is characterized by hyperglycemia in pregnancy and typically resolves after birth. Women with GDM have an increased risk of developing type 2 diabetes mellitus (T2DM) later in life compared to those with normoglycemic pregnancy. While diabetes prevention interventions (DPIs) have been developed to delay or prevent the onset of T2DM, few studies have provided process evaluation (PE) data to assess the mechanisms of impact, quality of implementation, or contextual factors that may influence the effectiveness of the intervention.

Objective: This study aims to identify and evaluate PE data and how these link to outcomes of randomized controlled trials (RCTs) of T2DM prevention interventions for women with GDM.

Methods: A systematic review was conducted to identify studies published from 2005 to 2020 aiming to capture the most recent DPIs. Five electronic bibliographic databases (Cochrane Library, Cochrane Collaboration Registry of Controlled Trials, Embase, PubMed, and MEDLINE) were searched to identify relevant studies. Inclusion criteria were published (peer-reviewed) RCTs of DPIs in women with a current diagnosis or history of GDM. Exclusion criteria were studies not published in English; studies where the target population was women who had a family history of T2D or women who were menopausal or postmenopausal; and gray literature, including abstracts in conference proceedings. The Medical Research Council’s PE framework of complex interventions was used to identify key PE components. The Mixed Method Appraisal Tool was used to assess the quality of included studies.

Results: A total of 24 studies were included; however, only 5 studies explicitly reported a PE theoretical framework. The studies involved 3 methods of intervention delivery, including in person (n=7), digital (n=7), and hybrid (n=9). Two of the studies conducted pilot RCTs assessing the feasibility and acceptability of their interventions, including recruitment, participation, retention, program implementation, adherence, and satisfaction, and 1 study assessed the efficacy of a questionnaire to promote food and vegetable intake. While most studies linked PE data with study outcomes, it was unclear which of the reported PE components were specifically linked to the positive outcomes.

XSL•FO
RenderX

Conclusions: While the Medical Research Council's framework is a valuable source for conducting systematic reviews on PEs, it has been criticized for lacking practical advice on how to conduct them. The lack of information on PE frameworks in our review also made it difficult to categorize individual PE components against the framework. We need clearer guidance and robust frameworks for conducting PEs for the development and reporting of DPIs for women with GDM.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020208212; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=208212

International Registered Report Identifier (IRRID): RR2-https://doi.org/10.1177/16094069211034010

(Interact J Med Res 2025;14:e51718) doi:10.2196/51718

KEYWORDS

gestational diabetes mellitus; randomized controlled trial; process evaluation; implementation; complex interventions

Introduction

Gestational diabetes mellitus (GDM) is characterized by hyperglycemia first recognized in pregnancy and currently impacts between 8% and 23% of pregnancies globally [1]. GDM typically resolves after birth but can have significant implications for both short- and long-term health of women and babies [2]. Women with prior GDM have a nearly 10-fold increased risk of developing type 2 diabetes mellitus (T2DM) than those with normoglycemic pregnancy [3]. Risk factors for T2DM after GDM include high BMI, increasing age, multiparity, poor glucose tolerance, and prepregnancy complications [4-6]. Lifestyle interventions that target both diet and exercise are associated with small but significant effects in reducing the risk of T2DM in women with GDM [7-10].

Process evaluation (PE) theoretical frameworks provide a systematic approach to planning the design of health behavior change interventions [11,12]. However, many of the randomized controlled trials (RCTs) reported not using these PE frameworks to plan the design of lifestyle interventions [11,13-17]. Several key PE theoretical frameworks have been developed and widely used [13-17], and researchers have made progress in updating the methodologies and definitions of key PE components to construct more comprehensive frameworks to measure the success and effectiveness of interventions [13]. Knowledge of these processes can better inform future policy and practice [18] and provide opportunities to improve study design and methodologies of future diabetes prevention interventions (DPIs) [19].

The UK Medical Research Council (MRC) theoretical framework on PEs for complex interventions was chosen for this systematic review [15], as it is well regarded and recognizes the need for more formal guidance on how to conduct PEs. The framework considers processes along 3 interlinked dimensions: implementation, mechanisms of impact, and contextual factors. Table 1 presents a glossary of key PE components and methodologies [15].

The overarching aim of this review was to identify and evaluate any PE conducted and how these link to outcomes of RCTs of DPIs after GDM. Specific objectives were to identify the extent to which PE components were reported and described in RCTs of DPIs for women with previous or current GDM and to assess whether these components could contribute to explaining the intervention outcomes of the DPIs.

XSL•FO
RenderX

Table 1. Glossary of key process evaluation components and methodologies.

Key process evaluation components	Definitions	Quantitative methods	Qualitative methods
Implementation			
Content delivered	The content included in the intervention and the method in which the content is delivered to the participants.	• Structured observation	• Observational study
Method of content delivery	How the intervention content was delivered to the participants and the extent to which quantitative and qualitative methods were used.	• Structured observation	• Observational study
Fidelity	The extent to which the intervention was delivered as intended. Fidelity represents the quality and integrity of the intervention.	• Behavioral coding systems • Questionnaires • Protocol checklist • Structured observation	• Audiotapes of sessions • Observational study
Dosage delivered	The quantity or number of intended units delivered during the intervention including the component delivered and the extent of the participant's engagement.	• Checklist records of dose delivered • Structured observation	• Audiotapes of sessions
Dosage received	The quantity or number of intended units received during the intervention including the component received by the participants.	• Behavioral coding systems • Questionnaires • Structured observation • Digital monitoring (eg, digital feedback)	• Audiotapes of sessions • Focus groups • Interviews • Observational monitoring
Adaptations	The extent to which alterations were made to an intervention to achieve better contextual fit.	• Structured observation	• Observational study
Reach	The proportion of the intended target audience that comes into contact with and participated in the intervention.	• Attendance lists • Standardized protocols	Observational monitoring
Mechanisms of impact			
Participant responses toward the intervention	The responses of participants who interacted with and received the intervention, their satisfaction, and the degree to which they found the intervention acceptable.	• Questionnaires	• Interviews • Focus groups
Mediators	The extent to which intermediary processes inform subsequent changes in outcomes.	• Structured observations	• Observational study
Unexpected pathways or consequences	The extent to which identifying unexpected pathways and mechanisms during the intervention meets the research needs and leads to intervention outcomes.	• Structured observations	• Observational study
Contextual factors			
Barriers	Any external factors that may act as a barrier toward the intervention implementation or its effect on the outcomes.	• Statistical analysis • Questionnaires	• Interviews • Focus groups • Observational study
Facilitators	Any external factors that may act as a facilitator to the intervention implementation or its effect on the outcomes.	• Statistical analysis • Questionnaires	• Interviews • Focus groups • Observational study

XSL•FO
RenderX

Methods

Study Design

This review used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines and was registered on PROSPERO (CRD42020208212). Owing to limited guidance on conducting systematic reviews in this area, we used the UK MRC theoretical framework from which we derived a glossary of components to use as a checklist to extract and categorize qualitative and quantitative data (Table 1) [15]. This framework considers 3 linked dimensions in which to consider processes: implementation, mechanisms of impact, and contextual factors. Implementation involves an assessment of fidelity, dose delivered and received (ie, how often or effectively it was delivered), adaptations made, reach, content delivered, and method of content delivery [20,21]. Mechanisms of impact involve an assessment of participants' satisfaction with the intervention, "unexpected consequences," and "mediators" [18]. Contextual factors refer to the external barriers and facilitators, such as cultural or organizational factors, that may alter the implementation of an intervention [22].

Search Strategy

The timeframe for the search was limited to between December 1, 2005, and December 16, 2020. This timeframe was selected because of the rapid increase in DPIs in addition to the fast-moving space of digital technology. The review aimed to capture the most recent DPIs. Five electronic bibliographic databases were used: Cochrane Library; Cochrane Collaboration Registry of Controlled Trials; Embase; PubMed; and MEDLINE. The main key search terms were GDM, RCT, and PE (see Multimedia Appendix 1 for a detailed list of search terms and strings and Multimedia Appendix 2 for key definitions). Reference lists of the included studies were also searched for additional eligible studies. Boolean search was used to combine the keywords with operators such as AND, NOT, and OR to further produce more relevant results, for example: ((GDM OR gestational diabetes OR pregnancy-induced diabetes OR diabetes in pregnancy) AND ((RCT OR controlled clinical trial OR pragmatic control trial OR clinical trial) AND (process evaluation OR program evaluation OR process assessment OR process acceptance OR outcome measures)); ((GDMs OR gestational diabetes OR pregnancy-induced diabetes)) AND ((RCT OR controlled clinical trial OR pragmatic clinical trial OR clinical trial) OR (process evaluation OR program assessment OR process acceptance OR outcome measures).

Eligibility Criteria

Inclusion criteria were published (peer-reviewed) RCTs of DPIs in women with a current diagnosis or history of GDM. Exclusion criteria were studies not published in English; studies where the target population was women who had a family history of T2D or women who were menopausal or postmenopausal; and gray literature, including abstracts in conference proceedings. For further details about the eligibility criteria, please see the study protocol [23].

Study Selection

After initial deduplication of the extracted data using Endnote (Clarivate), studies were uploaded to the platform Rayyan for screening. The 2 reviewers (IIMS and MB) independently screened the titles and abstracts of the studies. Disagreements were resolved via consensus or decided by a third reviewer (IPN) when consensus could not be reached. The full texts of potentially relevant studies were retrieved for further screening and independently appraised by 2 reviewers (IIMS and MB) for final inclusion.

Data Extraction, Study Characteristics, and Analysis

Standardized data extraction forms were developed and piloted. Extracted data included a summary of study characteristics, evaluation of processes (if any), methods, and findings. Data were extracted by a single reviewer (IIMS) and verified by a second reviewer (NH); discrepancies were again resolved via consensus.

We conducted a narrative synthesis of the studies as we anticipated that the heterogeneity of the methods used to assess processes would preclude meta-analysis, as has been found previously [10].

Assessment of Quality

The Mixed Method Appraisal Tool was used to assess the quality of included studies by 2 independent reviewers (IIMS) and (NH). The discrepancies were again resolved via consensus.

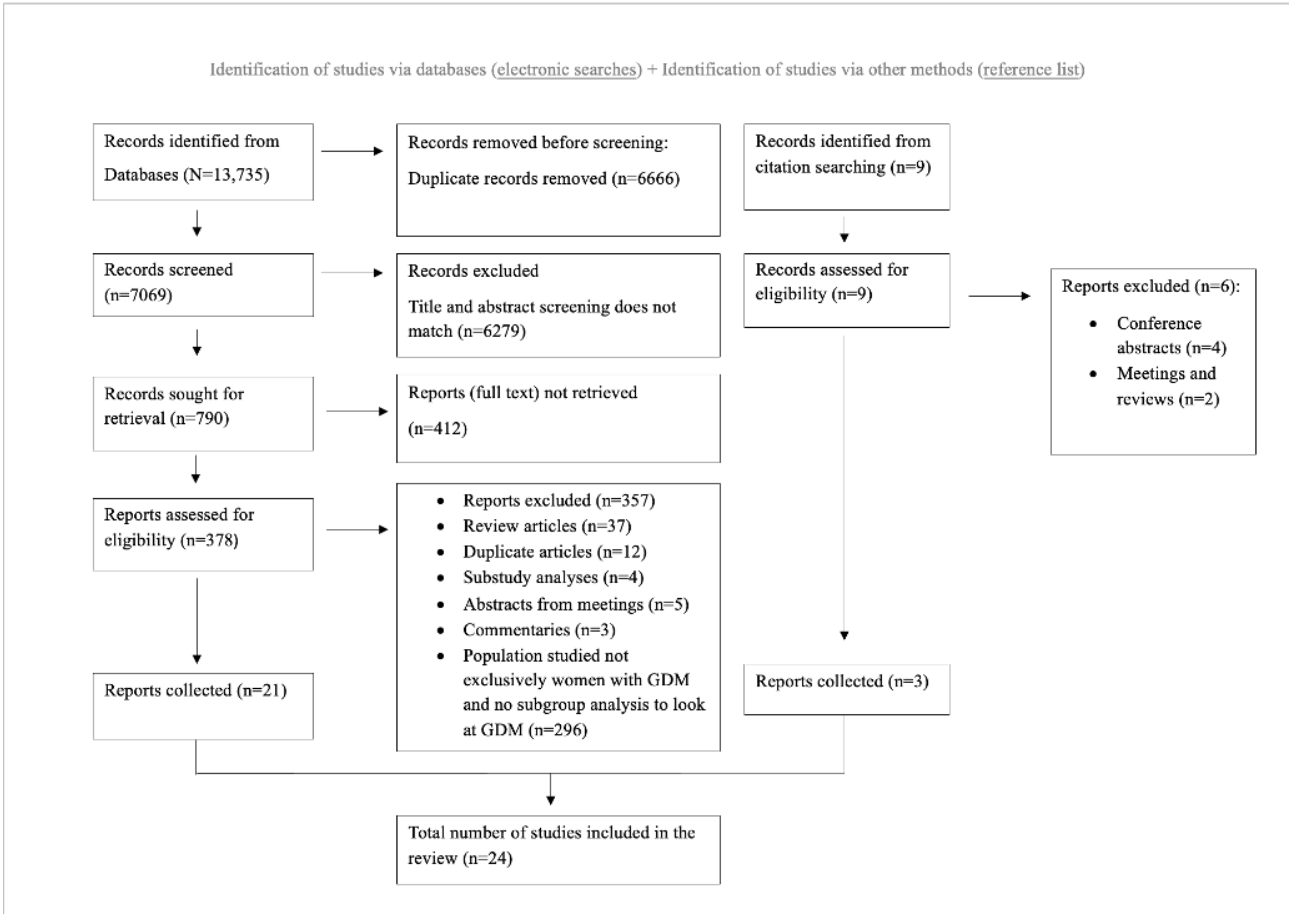
Results

Study Selection

The PRISMA flow diagram for this review is shown in Figure 1. Of the 13,735 records initially identified, the full text of 378 full-text studies were screened for eligibility and 21 studies met the eligibility criteria. An additional 3 studies were added following citation searching. Therefore, 24 studies were included in this review. Most of these studies fulfilled at least 3 methodological quality criteria outlined by the Mixed Method Appraisal Tool for each study design (Multimedia Appendix 3 [3,24-46]).

XSL•FO
RenderX

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram of search. GDM: gestational diabetes mellitus.



Study Characteristics

The stated primary aims of all 24 DPI studies reviewed were to improve dietary and physical activity outcomes, increase self-efficacy levels and risk perception of T2DM, and decrease maternal postpartum weight and BMI. The 24 studies evaluated in-person (n=8), digital (n=7), and hybrid (n=9) interventions. Two of the studies were pilot studies of the feasibility and acceptability of DPIs. All studies were published between 2007 and 2020. Most studies (n=22, 92%) were conducted in high-income countries. Only 2 studies were conducted in low- to middle-income countries: Egypt (n=1) and Malaysia (n=1). The sample size ranged from 31 to 1180 participants. The intervention duration varied from 12 weeks to 6 years. See Multimedia Appendix 4 [3,24-46] for a description of study characteristics of the studies included in this review.

Number and Type of PE Components Reported

The most frequent PE components evaluated were content delivered and method of content delivery (n=24), dose delivered (n=20), reach (n=20), dose received (n=17), barriers (n=15), facilitators (n=8), and participants' responses toward the intervention (n=8). Relatively few studies reported process data relating to adaptations (n=7), mediators (n=7), fidelity (n=5), or unexpected pathways or consequences (n=2). All PE components outlined by the MRC framework and reported in the reviewed DPIs are summarized in Multimedia Appendices 5 and 6 [3,24-46].

PE Frameworks and Measures Reported

Five studies explicitly referred to a PE framework or PE quantitative or qualitative measures. Only 1 of these referred to process measures in the context of the MRC framework of complex interventions [24]. The others reported measures to ensure adherence to the study protocol [25], to measure "reach" (using the penetration, implementation, participation, and effectiveness metric) [26,27], or to assess participants' satisfaction to the program using a survey [47].

Implementation

Content Delivered and Method of Content Delivery

Eight studies evaluated in-person or group DPIs [3,28-34], 7 evaluated digital DPIs [35-40,48], and 9 delivered hybrid DPIs [24-27,41-44]. Most of the interventions were multicomponent that focused on diet and self-directed physical activity [24-27,33,35,40-46,49] or were facilitator led [29,30,43]. One intervention also included advice on breastfeeding [28].

Behavior Techniques and Theoretical Frameworks Used to Deliver Content

Behavior change techniques included motivational interviewing (MI) with individual goal setting [26,29]; enhancing healthy lifestyle change and self-management [24]; guided self-help [33]; nutritional and physical activity recommendations [30]; self-help guidance based on responses to a dietary food intake questionnaire [3]; and guideline-based nutritional strategies [44]. Five interventions were underpinned by established

XSL•FO
RenderX

theoretical frameworks—social cognitive theory and the transtheoretical model of behavior change [25]; the Health Action Process Approach supported by social cognitive and self-regulation theory [27] based on a Health Belief Model [50]; and messages on weight loss, diet, and physical activity based on a harm reduction model to promote healthy lifestyle changes [43].

The link between a process and outcome was reported in 2 studies. In the first, the use of behavior change strategies was reported to improve women's knowledge, beliefs, and self-reported practices as well as decrease postpartum weight gain in 1 study [33]. In the second study, there were no observed changes in weight, waist circumference, or blood glucose level and this was correlated with no changes in health perception or self-efficacy [3].

Fidelity

Five DPIs (digital: n=3 and hybrid: n=2) reported on the fidelity (quality) of intervention implementation [25,27,36,39,45]. Assessment of fidelity involved recording telephone calls or sessions [36,45], a qualitative [45] or quantitative [36,39] assessment of adherence, and sometimes also involved input from the participants [45]. Two of the studies attributed a high level of fidelity to effective weight loss and weight goals [27,36].

Dose Delivered

Nineteen studies reported on the quantity, frequency, and duration of the interventions [24-26,28-30,32,34,36,39-43,48]. Most studies delivered around 1 to 4 sessions lasting 2 hours. The shortest dose duration was 12 weeks and included 12-weekly sessions of 2.5 hours including group physical activity, health education, and an individualized lifestyle counseling session [29].

Further details in relation to dose delivery were generally limited. For instance, the frequency [24,40] or quantity of sessions or emails [43,45] were not reported. Similarly, 3 digital DPIs of web-based interventions involving health websites [40], internet telemedicine [37], or pedometer messaging [38] did not report the quantity of health content and messages sent by health care professionals. Borgen et al [35] reported the delivery of their mobile health app but did not mention how long it would take to peruse the app's content. Finally, another web-based intervention [44] did not report on how frequently participants should use the pedometer or step count goal target.

Dose Received

Sixteen of the studies reported data on the dose received by the participants [3,24-26,28-30,32,36-38,43,45]. The dose response included the average number of sessions attended by women [25-27,29,36,40,43-45]; those returning documents or questionnaires [51-53]; combined food intake and physical activity logbooks [38,54,55]; and log-ins to websites or systems [37,40,45].

Information on the dose received was used to highlight the willingness of participants to take part in the study and the subsequent success of the intervention. For example, in 1 hybrid study [44], it was reported that most women in the study attended all 4 lifestyle counseling sessions in their intervention, concluding that it was successful in reducing weight and increasing physical activity levels. Details on the dose received were also used to explain any lack of intervention effects. In another hybrid study [43], authors attributed their unsuccessful intervention outcomes to the limited number of women logging onto their website intervention and observed that engagement did not improve with fortnightly automated email reminders. One digital study observed that they did not collect data on user engagement for its mobile health app to protect women's privacy and questioned that this may have explained the nonsignificant effect of the intervention [35].

Adaptations

Adaptations were described in 7 DPIs, including 2 in-person [38,55], 2 digital [35,45], and 3 hybrid DPIs [24,25,32]. These adaptations were seen as important components for the success of the intervention and included providing a choice of intervention methods [28] and tailoring interventions to the characteristics of the target population [45] to specific ethnic or cultural backgrounds [24,34,35]. In 3 hybrid DPIs, MI techniques were used to tailor sessions for participants [24,25] and illustrations and simple messages were added to address low health literacy [41].

Reach

Twenty studies reported on the reach of its intervention in relation to the target population [24-26,28-32,34,36-41,43,45]. Loss to follow-up was a recognized challenge, perhaps reflecting the case mix of the target population. Two of the studies referred to pregnancy complications as a reason for dropout [31,33] with another attributing loss to follow-up to work or personal commitments, initiation of a weight loss diet, subsequent pregnancy, not being contactable, and finding intervention resources unhelpful [40]. One digital DPI reported a low participation rate of 17%, making inference to the broader target population difficult [39]; another reported that few women who received information about their intervention proceeded to enroll for participation [38]. Other studies that used PE methods to measure reach reported low rates [24,27,43] and 2 additional studies had 33% to 39% attrition rates [29,34].

Mechanisms of Impact

Participant Responses Toward the Intervention

Eight studies reported on participant feedback and responses toward the interventions including 4 digital [35,38-40], 3 hybrid [25,26,44], and 1 in-person DPI [29]. Overall, women expressed satisfaction with the interventions [25,29,38,39,44], reporting increased confidence in setting health goals [39,40] and more engagement with health management [35].

Some studies did not receive a positive response. Only a third of women reported being motivated by the website content in 1 digital intervention [40] and even fewer (22% and 31%) had a positive reaction to the text messaging component. Women also criticized the lack of information on optimal carbohydrates to consume when transitioning between pregnancy to postpartum diets and reported the need for more low-fat recipes [25,40].

XSL•FO
RenderX

Mediators

Mediators were described in the interventions in 2 in-person DPI studies [29,34], 4 digital DPIs [35,37,39,40], and 1 hybrid DPI [43], although not explicitly. O'Dea et al [29] found that women valued individual face-to-face sessions with health care professionals when setting health goals which improved their stress levels, diet self-efficacy, and quality of life. Social relationships also developed between the women due to regular attendance at the same lifestyle counseling groups [34]. Digital DPIs were also reported to increase women's engagement in their own health [35], enhance self-efficacy and confidence [37,40], and provide reassurance in developing and attaining health goals [39]. Further, 1 hybrid study [43] qualitatively linked the inclusion of partners in their intervention as a key mediator for women making them more likely to participate in and engage with the intervention.

Unexpected Pathways or Consequences

Only 2 studies [39,41] reported unexpected consequences in relation to their interventions. Carolan-Olah and Sayakhot [41] observed there was an unusually high percentage of women in the intervention group who had attended their postpartum oral glucose tolerance test (OGTT) appointment. Although not directly linked to the study, the intervention highlighted the importance of postpartum testing, which may have increased women's motivation to attend future follow-up appointments. Another digital study [39] experienced unexpected challenges with data retrieval from women and general practitioners, resulting in missing baseline blood tests and self-reported data at follow-up. However, this missing data did not appear to negatively influence the findings as women in the intervention group were found to significantly reduce their fat intake.

Contextual Factors

Barriers

Fifteen studies referred to barriers that women faced when participating in interventions [3,24-26,29,31,33-35,37-40,43,44]. Such barriers included lack of time due to family and work commitments, subsequent pregnancy, and changes to daily life due to the demands of motherhood [24,29,31,33,40,44] technology issues, and lack of internet access [24,38,43]. Similarly, unavailability due to childcare responsibilities and forgetting to attend study visits prevented mothers from engaging in in-person lifestyle counseling sessions in the nondigital interventions [3,24,29]. Another in-person DPI acknowledged how culture plays a role in hindering women from participating in and completing an intervention when they are placed under pressure to commit to family and social norms [34].

Facilitators

The 8 DPI studies including 4 in-person DPIs [28,29,31,34], 2 hybrid DPIs [43,44], and 2 digital DPIs [37,40] described facilitators as contextual factors which may have influenced the outcomes of their interventions. The most common facilitator identified in 1 hybrid and 1 in-person DPI was receiving support from a partner during the intervention [29,43] and having higher levels of income and personal education background [43]. Having access to healthy food, food vouchers, a babysitter, and an exercise buddy were also reported as facilitators to assist and support the uptake and maintenance of postpartum healthy behaviors [31,40]. The competency and skills of health care professionals delivering the interventions were acknowledged as facilitators to ensure the uptake of the intervention [28]. Having an accessible, central location for assessments and employers who allowed time for visits also facilitated participation [34,44].

Discussion

Overview

We identified and reviewed 24 RCTS of DPIs that evaluated at least 1 PE component for women with previous or current GDM. Only 10 studies explicitly reported individual processes, and only 5 explicitly referred to a PE framework or PE quantitative or qualitative measures to assess processes. Overall, few studies reported and evaluated processes in relation to study outcomes, and although most DPIs linked PE components and DPI outcomes in some capacity, it was challenging to attribute any 1 process component to the effectiveness and success of the intervention.

Summary of Key Findings

The most complete of the PE components reported across all 24 studies was the combined component of content delivered and method of content delivery. Clear narratives were provided on how the interventions were delivered and their content. Most of the interventions also referred to broad behavior change techniques such as goal setting and MI, which are known to be effective in promoting lifestyle behavior change [56,57]. However, few studies linked these theories with mechanisms of action by providing a detailed psychological or behavior change theoretical framework on which their intervention was underpinned. Such frameworks are important to make sense of "how," "why," and "under what circumstances" intervention components work together to achieve the desired outcomes. Moving forward, we recommend that all RCTs of DPIs publish an overview of the theories they have used to design and evaluate the intervention for greater transparency.

Although most DPIs reported information on the dose delivered, PE measures on the dose received were less complete. Generally, increased frequency and longer sessions resulted in more engagement with the intervention and better outcomes [26-28,30,31,36]. More frequent sessions may also be preferred by the participants: for a study on the use of MI and hemoglobin A_{1c} outcomes, women said that they would like more sessions to improve their engagement with the intervention [58]. The timing of the intervention is also a consideration. Most of the DPIs that we reviewed delivered interventions during either the antenatal or postpartum periods. However, systematic review evidence suggests that optimal pregnancy outcomes can be achieved by delivering the intervention during both periods since engagement with a healthy lifestyle poses tends to be poorer during the postpartum stage [5,51]. Moreover, there was little information in the studies reviewed on how the dose received impacted the success of the intervention. The challenges in assessing adherence to the dose received are well

XSL•FO
RenderX

recognized as people may have received a dose but choose not to take it [59]. This is also supported in the literature where a nurse-led psychological intervention for a T2D cluster RCT study identified a lower dose received than intended, but that this was not associated with the dose delivered [60]. Further research to explore how timing, frequency, and duration of the intervention affect outcomes among women with current or previous GDM is indicated.

Although all 24 studies included in this review sought to describe the reach of the intervention, few gave a detailed evaluation of this measure. Poor reach was attributed to generic reasons, most commonly low participation rates, loss to follow-up, and small sample sizes [26,27,29,36,39]. The increase in OGTT appointment attendance found in 1 study [41] is contrary to prior research linking loss to follow-up and withdrawal numbers to participants' reluctance to attend OGTT appointments [49,54], which highlights the influence that different contexts have on outcomes in this population. Moreover, the studies that made adaptations to tailor their DPIs to personal or cultural needs [24,25,28,34,35,41,45] are likely to have increased reach by making the interventions more relevant. A clear definition of the target population and consideration of how the DPI can be adapted to make it as inclusive as possible seems important for its success.

Most of the studies reported on barriers to intervention engagement. Similar to previous systematic reviews [10,61], the barriers primarily related to lack of time, childcare commitments, and challenges to maintaining healthy lifestyles [24,29]. Focus groups with women suggest that these inherent barriers can be mitigated by combining direct contact with health care professionals with web-based interventions [48]. Moreover, it is worth noting that, despite the many barriers to participation identified, women were still in favor of the DPIs, and many resulted in positive outcomes. This aligns with previous research in T2D [62] and supports a need and willingness among women with previous or current GDM to engage with DPIs.

A particular concern identified by this review was the lack of reporting or evaluation of fidelity. Only 5 studies reported on this process component. Ensuring fidelity is a key aspect of any DPI because it safeguards against nonadherence to the study protocols and inadequate implementation delivery [47,63]. Ideally, fidelity should be measured prior to conducting full-scale implementation to distinguish between outcomes that are related to ineffectiveness from those related to protocol deviation [64,65]. In our review, methods to assess fidelity included audiotaping and patient registers of attendance [27,36]. Such methods have been tried and tested in the T2D setting [47,63] and are likely to be more robust than relying on the expertise of health care professionals as a fidelity measure [66]. To minimize research waste, fidelity assessments should be commenced at the design stage of a trial and incorporate standardized measures where possible.

Finally, we welcome the use of both quantitative and qualitative methods to evaluate the PE components in our studies because this enables a more in-depth understanding of the relationship between individual components and outcomes [52,67]. We reemphasize the importance of mixed methods when conducting thorough PE.

Strengths and Limitations

To the best of our knowledge, this is the first and only systematic review of PEs of T2DM prevention interventions for women with GDM. We used robust methodology, an established framework of PE for complex interventions [15], and restricted the review to RCTs, which are known to be the best design for evaluating complex health care interventions [15]. The glossary list of standardized DPI content and PE terms developed during the review process can be used to guide future RCTs in this area. However, while the MRC framework [18] is a valuable source for conducting systematic reviews on PEs, it has been criticized for lacking practical advice on how to conduct them [68]. The lack of information on PE frameworks in our review also made it difficult to categorize individual PE components against the framework. Moreover, there is a lack of consensus on how PE components should be defined and interpreted by researchers, which is only partially addressed in the MRC guidelines [15], which facilitate understanding of PE constructs but provide little clarity on what should be measured and how.

Conclusions

This systematic review has highlighted that there are important gaps in the reporting of PE metrics for RCTs of T2DM in women with GDM. We recommend rigorous, systematic, and in-depth PE guidance to facilitate reporting of these studies. Future research should focus on reaching consensus on the reporting of PE measures using established frameworks and evaluating PE in real-world health care settings to optimize the interpretation of study outcomes.

Acknowledgments

This work is supported by MY PAIR Grant UK-Malaysia: Joint Partnership Call on Non-Communicable Diseases (Malaysia: JPT.S(BPKI)2000/011/06/05 (27); United Kingdom: MR/T018240/1). This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London and Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care.

The authors confirm that generative artificial intelligence was not used to conduct or report any portion of this systematic review.

XSL•FO
RenderX

Data Availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files).

Authors' Contributions

IIMS developed the methodology, conducted the investigation and formal analysis, and cowrote and reviewed the original draft. MB conducted the investigation, cowrote the original draft, reviewed and edited subsequent drafts, supervised junior researchers, and contributed to project administration. IPN codeveloped the original concept, methodology, and investigation; curated data and cowrote the original draft; edited subsequent drafts; supervised junior researchers; and acquired funding. AB cowrote the original draft and reviewed subsequent drafts. MP developed the original methodology, contributed to the investigation, cowrote drafts, and supervised junior researchers. BNMY contributed to the writing of drafts and reviewing subsequent versions. SB contributed to writing, reviewed, and edited subsequent versions. CSM codeveloped the original concept, contributed to the writing and reviewing of subsequent versions, and supervised junior researchers. AF codeveloped the original concept, developed the methodology, cowrote and reviewed subsequent versions of the original draft, and supervised junior researchers. KI codeveloped the original concept; developed the methodology; conducted the formal analysis; contributed to the investigation, data curation, and inputting of data into the software; supervised junior researchers; contributed to project administration; and acquired funding. BHC codeveloped the original concept, developed the methodology, cowrote the original draft and reviewed subsequent versions, supervised junior researchers, and acquired funding. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Detailed list of search terms and strings.
[DOCX File , 16 KB - ijmr_v14i1e51718_app1.docx]

Multimedia Appendix 2
Key definitions of types and formats of DPI content delivered, behavior change techniques, and psychological theoretical frameworks embedded within DPIs. DPI: diabetes prevention intervention.
[DOCX File , 17 KB - ijmr_v14i1e51718_app2.docx]

Multimedia Appendix 3
Quality assessment of review articles.
[DOCX File , 26 KB - ijmr_v14i1e51718_app3.docx]

Multimedia Appendix 4
Study characteristics of interventions for the prevention of type 2 diabetes in women with gestational diabetes mellitus.
[DOCX File , 47 KB - ijmr_v14i1e51718_app4.docx]

Multimedia Appendix 5
Process evaluation implementation components for T2DM prevention lifestyle intervention studies on women with GDM. GDM: gestational diabetes mellitus; T2DM: type 2 diabetes mellitus.
[DOCX File , 31 KB - ijmr_v14i1e51718_app5.docx]

Multimedia Appendix 6 [DOCX File , 21 KB - ijmr_v14i1e51718_app6.docx]

Multimedia Appendix 7
PRISMA Checklist for systematic review.
[PDF File (Adobe PDF File), 58 KB - ijmr_v14i1e51718_app7.pdf]

References

1. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, et al. IDF Diabetes Atlas: Global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract 2018 Apr;138:271-281. [doi: 10.1016/j.diabres.2018.02.023] [Medline: 29496507]
2. Kytö M, Koivusalo S, Tuomonen H, Strömberg L, Ruonala A, Marttinen P, et al. Supporting the management of gestational diabetes mellitus with comprehensive self-tracking: mixed methods study of wearable sensors. JMIR Diabetes 2023 Oct 31;8:e43979 [FREE Full text] [doi: 10.2196/43979] [Medline: 37906216]

XSL•FO
RenderX

3. Vézina-Im LA, Perron J, Lemieux S, Robitaille J. Promoting fruit and vegetable intake in childbearing age women at risk for gestational diabetes mellitus: A randomised controlled trial. J Health Psychol 2019 Apr;24(5):600-612. [doi: 10.1177/1359105316680021] [Medline: 27872389]
4. Baptiste-Roberts K, Barone BB, Gary TL, Golden SH, Wilson LM, Bass EB, et al. Risk factors for type 2 diabetes among women with gestational diabetes: a systematic review. Am J Med 2009 Mar;122(3):207-214.e4 [FREE Full text] [doi: 10.1016/j.amjmed.2008.09.034] [Medline: 19272478]
5. Li Z, Cheng Y, Wang D, Chen H, Chen H, Ming WK, et al. Incidence rate of type 2 diabetes mellitus after gestational diabetes mellitus: a systematic review and meta-analysis of 170,139 women. J Diabetes Res 2020;2020:3076463 [FREE Full text] [doi: 10.1155/2020/3076463] [Medline: 32405502]
6. Rayanagoudar G, Hashi AA, Zamora J, Khan KS, Hitman GA, Thangaratinam S. Quantification of the type 2 diabetes risk in women with gestational diabetes: a systematic review and meta-analysis of 95,750 women. Diabetologia 2016;59(7):1403-1411 [FREE Full text] [doi: 10.1007/s00125-016-3927-2] [Medline: 27073002]
7. Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: an exposure-wide umbrella review of meta-analyses. PLoS One 2018;13(3):e0194127 [FREE Full text] [doi: 10.1371/journal.pone.0194127] [Medline: 29558518]
8. Brown J, Alwan NA, West J, Brown S, McKinlay CJ, Farrar D, et al. Lifestyle interventions for the treatment of women with gestational diabetes. Cochrane Database Syst Rev 2017;5(5):CD011970 [FREE Full text] [doi: 10.1002/14651858.CD011970.pub2] [Medline: 28472859]
9. Gilinsky AS, Kirk AF, Hughes AR, Lindsay RS. Lifestyle interventions for type 2 diabetes prevention in women with prior gestational diabetes: a systematic review and meta-analysis of behavioural, anthropometric and metabolic outcomes. Prev Med Rep 2015;2:448-461 [FREE Full text] [doi: 10.1016/j.pmedr.2015.05.009] [Medline: 26844102]
10. Guo J, Chen JL, Whittemore R, Whitaker E. Postpartum lifestyle interventions to prevent type 2 diabetes among women with history of gestational diabetes: a systematic review of randomized clinical trials. J Womens Health 2016;25(1):38-49. [doi: 10.1089/jwh.2015.5262] [Medline: 26700931]
11. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
12. Fernandez ME, Ruiter RAC, Markham CM, Kok G. Intervention mapping: theory- and evidence-based health promotion program planning: perspective and examples. Front Public Health 2019;7:209 [FREE Full text] [doi: 10.3389/fpubh.2019.00209] [Medline: 31475126]
13. Baranowski T, Stables G. Process evaluations of the 5-a-day projects. Health Educ Behav 2000;27(2):157-166. [doi: 10.1177/109019810002700202] [Medline: 10768797]
14. Grant A, Treweek S, Dreischulte T, Foy R, Guthrie B. Process evaluations for cluster-randomised trials of complex interventions: a proposed framework for design and reporting. Trials 2013;14:15 [FREE Full text] [doi: 10.1186/1745-6215-14-15] [Medline: 23311722]
15. Moore GF, Audrey S, Barker M, Bond L, Bonell C, Hardeman W, et al. Process evaluation of complex interventions: medical research council guidance. BMJ 2015;350:h1258 [FREE Full text] [doi: 10.1136/bmj.h1258] [Medline: 25791983]
16. Saunders RP, Evans MH, Joshi P. Developing a process-evaluation plan for assessing health promotion program implementation: a how-to guide. Health Promot Pract 2005;6(2):134-147. [doi: 10.1177/1524839904273387] [Medline: 15855283]
17. Steckler AE, Linnan LE. Process Evaluation for Public Health Interventions and Research. San Francisco, CA: Jossey-Bass; 2002.
18. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, et al. Developing and evaluating complex interventions: the new medical research council guidance. BMJ 2008;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
19. Whittemore R, Grey M. The systematic development of nursing interventions. J Nurs Scholarsh 2002;34(2):115-120. [doi: 10.1111/j.1547-5069.2002.00115.x] [Medline: 12078534]
20. Carroll C, Patterson M, Wood S, Booth A, Rick J, Balain S. A conceptual framework for implementation fidelity. Implement Sci 2007;2:40 [FREE Full text] [doi: 10.1186/1748-5908-2-40] [Medline: 18053122]
21. Montgomery P, Underhill K, Gardner F, Operario D, Mayo-Wilson E. The Oxford Implementation Index: a new tool for incorporating implementation data into systematic reviews and meta-analyses. J Clin Epidemiol 2013;66(8):874-882 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.03.006] [Medline: 23810026]
22. Durlak JA, DuPre EP. Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. Am J Community Psychol 2008;41:327-350. [doi: 10.1007/s10464-008-9165-0] [Medline: 18322790]
23. Mohd Sa'id II, Papachristou Nadal I, Forbes A, Goldsmith K, Ismail IZ, Hassan F, et al. A protocol of process evaluations of interventions for the prevention of type 2 diabetes in women with gestational diabetes mellitus: a systematic review. Int J Qual Methods 2021;20:160940692110340. [doi: 10.1177/16094069211034010]

XSL•FO
RenderX

24. Holmes VA, Draffin CR, Patterson CC, Francis L, Irwin J, McConnell M, et al. Postnatal lifestyle intervention for overweight women with previous gestational diabetes: a randomized controlled trial. J Clin Endocrinol Metab 2018;103(7):2478-2487. [doi: 10.1210/jc.2017-02654] [Medline: 29762737]
25. Ferrara A, Hedderson MM, Albright CL, Ehrlich SF, Quesenberry CP, Peng T, et al. A pregnancy and postpartum lifestyle intervention in women with gestational diabetes mellitus reduces diabetes risk factors: a feasibility randomized control trial. Diabetes Care 2011;34(7):1519-1525 [FREE Full text] [doi: 10.2337/dc10-2221] [Medline: 21540430]
26. Lipscombe LL, Delos-Reyes F, Glenn AJ, de Sequeira S, Liang X, Grant S, et al. The avoiding diabetes after pregnancy trial in moms program: feasibility of a diabetes prevention program for women with recent gestational diabetes mellitus. Can J Diabetes 2019;43(8):613-620. [doi: 10.1016/j.jcjd.2019.08.019] [Medline: 31669188]
27. O'Reilly SL, Dunbar JA, Versace V, Janus E, Best JD, Carter R, et al. Mothers After Gestational Diabetes in Australia (MAGDA): a randomised controlled trial of a postnatal diabetes prevention program. PLoS Med 2016;13(7):e1002092 [FREE Full text] [doi: 10.1371/journal.pmed.1002092] [Medline: 27459502]
28. Koivusalo SB, Rönö K, Klemetti MK, Roine RP, Lindström J, Erkkola M, et al. Gestational diabetes mellitus can be prevented by lifestyle intervention: the Finnish Gestational Diabetes Prevention Study (RADIEL): a randomised controlled trial. Diabetes Care 2016;39(1):24-30. [doi: 10.2337/dc15-0511] [Medline: 26223239]
29. O'Dea A, Tierney M, McGuire BE, Newell J, Glynn LG, Gibson I, et al. Can the onset of type 2 diabetes be delayed by a group-based lifestyle intervention in women with prediabetes following gestational diabetes mellitus (GDM)? Findings from a randomized control mixed methods trial. J Diabetes Res 2015;2015:798460 [FREE Full text] [doi: 10.1155/2015/798460] [Medline: 26347894]
30. Pérez-Ferre N, Del Valle L, Torrejón MJ, Barca I, Calvo MI, Matía P, et al. Diabetes mellitus and abnormal glucose tolerance development after gestational diabetes: a three-year, prospective, randomized, clinical-based, mediterranean lifestyle interventional study with parallel groups. Clin Nutr 2015;34(4):579-585. [doi: 10.1016/j.clnu.2014.09.005] [Medline: 25262459]
31. Shek NWM, Ngai CSW, Lee CP, Chan JYC, Lao TTH. Lifestyle modifications in the development of diabetes mellitus and metabolic syndrome in Chinese women who had gestational diabetes mellitus: a randomized interventional trial. Arch Gynecol Obstet 2014;289(2):319-327. [doi: 10.1007/s00404-013-2971-0] [Medline: 23897066]
32. Shyam S, Arshad F, Abdul Ghani R, Wahab NA, Safii NS, Nisak MYB, et al. Low glycaemic index diets improve glucose tolerance and body weight in women with previous history of gestational diabetes: a six months randomized trial. Nutr J 2013;12(1):1-12 [FREE Full text] [doi: 10.1186/1475-2891-12-68] [Medline: 23705645]
33. Tawfik MY. The impact of health education intervention for prevention and early detection of type 2 diabetes in women with gestational diabetes. J Community Health 2017;42(3):500-510. [doi: 10.1007/s10900-016-0282-7] [Medline: 27743337]
34. Zilberman-Kravits D, Meyerstein N, Abu-Rabia Y, Wiznitzer A, Harman-Boehm I. The impact of a cultural lifestyle intervention on metabolic parameters after gestational diabetes mellitus a randomized controlled trial. Matern Child Health J 2018;22(6):803-811. [doi: 10.1007/s10995-018-2450-0] [Medline: 29411251]
35. Borgen I, Småstuen MC, Jacobsen AF, Garnweidner-Holme LM, Fayyad S, Noll J, et al. Effect of the Pregnant+ smartphone application in women with gestational diabetes mellitus: a randomised controlled trial in norway. BMJ Open 2019;9(11):e030884 [FREE Full text] [doi: 10.1136/bmjopen-2019-030884] [Medline: 31719080]
36. Ferrara A, Hedderson MM, Brown SD, Albright CL, Ehrlich SF, Tsai AL, et al. The comparative effectiveness of diabetes prevention strategies to reduce postpartum weight retention in women with gestational diabetes mellitus: the Gestational Diabetes' Effects on Moms (GEM) cluster randomized controlled trial. Diabetes Care 2016;39(1):65-74 [FREE Full text] [doi: 10.2337/dc15-1254] [Medline: 26657945]
37. Homko CJ, Santamore WP, Whiteman V, Bower M, Berger P, Geifman-Holtzman O, et al. Use of an internet-based telemedicine system to manage underserved women with gestational diabetes mellitus. Diabetes Technol Ther 2007;9(3):297-306. [doi: 10.1089/dia.2006.0034] [Medline: 17561800]
38. Kim C, Draska M, Hess ML, Wilson EJ, Richardson CR. A web-based pedometer programme in women with a recent history of gestational diabetes. Diabet Med 2012;29(2):278-283 [FREE Full text] [doi: 10.1111/j.1464-5491.2011.03415.x] [Medline: 21838764]
39. Reinhardt JA, van der Ploeg HP, Grzegrzulka R, Timperley JG. lmplementing lifestyle change through phone-based motivational interviewing in rural-based women with previous gestational diabetes mellitus. Health Promot J Austr 2012;23(1):5-9. [doi: 10.1071/he12005] [Medline: 22730940]
40. Rollo ME, Baldwin JN, Hutchesson M, Aguiar EJ, Wynne K, Young A, et al. The feasibility and preliminary efficacy of an eHealth lifestyle program in women with recent gestational diabetes mellitus: a pilot study. Int J Environ Res Public Health 2020;17(19):7115. [doi: 10.3390/ijerph17197115] [Medline: 32998401]
41. Carolan-Olah M, Sayakhot P. A randomized controlled trial of a web-based education intervention for women with gestational diabetes mellitus. Midwifery 2019;68:39-47. [doi: 10.1016/j.midw.2018.08.019] [Medline: 30343264]
42. Hu G, Tian H, Zhang F, Liu H, Zhang C, Zhang S, et al. Tianjin Gestational Diabetes Mellitus Prevention Program: study design, methods, and 1-year interim report on the feasibility of lifestyle intervention program. Diabetes Res Clin Pract 2012;98(3):508-517. [doi: 10.1016/j.diabres.2012.09.015] [Medline: 23010556]

XSL•FO
RenderX

43. McManus R, Miller D, Mottola M, Giroux I, Donovan L. Translating healthy living messages to postpartum women and their partners after gestational diabetes (GDM): body habitus, A1C, lifestyle habits, and program engagement results from the families defeating diabetes (FDD) randomized trial. Am J Health Promot 2018;32(6):1438-1446. [doi: 10.1177/0890117117738210] [Medline: 29108443]
44. Peacock AS, Bogossian FE, Wilkinson SA, Gibbons KS, Kim C, McIntyre HD. A randomised controlled trial to delay or prevent type 2 diabetes after gestational diabetes: walking for exercise and nutrition to prevent diabetes for you. Int J Endocrinol 2015;2015:423717 [FREE Full text] [doi: 10.1155/2015/423717] [Medline: 26089886]
45. Nicklas JM, Zera CA, England LJ, Rosner BA, Horton E, Levkoff SE, et al. A web-based lifestyle intervention for women with recent gestational diabetes mellitus: a randomized controlled trial. Obstet Gynecol 2014;124(3):563-570 [FREE Full text] [doi: 10.1097/AOG.0000000000000420] [Medline: 25162257]
46. Liu H, Wang L, Zhang S, Leng J, Li N, Li W, et al. One-year weight losses in the Tianjin Gestational Diabetes Mellitus Prevention Programme: a randomized clinical trial. Diabetes Obes Metab 2018;20(5):1246-1255 [FREE Full text] [doi: 10.1111/dom.13225] [Medline: 29360237]
47. West DS, DiLillo V, Bursac Z, Gore SA, Greene PG. Motivational interviewing improves weight loss in women with type 2 diabetes. Diabetes Care 2007;30(5):1081-1087. [doi: 10.2337/dc06-1966] [Medline: 17337504]
48. Nicklas JM, Zera CA, Seely EW, Abdul-Rahim ZS, Rudloff ND, Levkoff SE. Identifying postpartum intervention approaches to prevent type 2 diabetes in women with a history of gestational diabetes. BMC Pregnancy Childbirth 2011;11:23 [FREE Full text] [doi: 10.1186/1471-2393-11-23] [Medline: 21435246]
49. Kim C, Newton KM, Knopp RH. Gestational diabetes and the incidence of type 2 diabetes: a systematic review. Diabetes Care 2002;25(10):1862-1868. [doi: 10.2337/diacare.25.10.1862] [Medline: 12351492]
50. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
51. Huang S, Magny-Normilus C, McMahon E, Whittemore R. Systematic review of lifestyle interventions for gestational diabetes mellitus in pregnancy and the postpartum period. J Obstet Gynecol Neonatal Nurs 2022;51(2):115-125 [FREE Full text] [doi: 10.1016/j.jogn.2021.10.007] [Medline: 34843670]
52. O'Cathain A, Murphy E, Nicholl J. Why, and how, mixed methods research is undertaken in health services research in England: a mixed methods study. BMC Health Serv Res 2007;7:85 [FREE Full text] [doi: 10.1186/1472-6963-7-85] [Medline: 17570838]
53. Gollwitzer PM, Oettingen G. Implementation intentions. In: Encyclopedia of Behavioural Medicine. Cham, Switzerland: Springer International Publishing; 2020:1159-1164.
54. Buckley BS, Harreiter J, Damm P, Corcoy R, Chico A, Simmons D, et al. Gestational diabetes mellitus in Europe: prevalence, current screening practice and barriers to screening. A review. Diabet Med 2012;29(7):844-854. [doi: 10.1111/j.1464-5491.2011.03541.x] [Medline: 22150506]
55. Schunk DH, Usher EL. Social cognitive theory and motivation. In: Ryan RM, editor. The Oxford Handbook of Human Motivation. New York, NY: Oxford University Press; 2012:13-27.
56. Michie S. Designing and implementing behaviour change interventions to improve population health. J Health Serv Res Policy 2008;13:64-69. [doi: 10.1258/jhsrp.2008.008014] [Medline: 18806194]
57. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, 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;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
58. Waker CL. Effects of motivational interviewing on diabetes self-management behaviors and glycemic control in type 2 diabetes: a translational study [dissertation]. OH: University of Cincinnati; 2012. URL: http://rave.ohiolink.edu/etdc/view?acc_num=ucin1342728810 [accessed 2024-08-26]
59. Graves H, Ismail K, Winkley K. A framework for process evaluations of psychological interventions [Dissertation]. King's College London. 2020. URL: https://kclpure.kcl.ac.uk/portal/en/studentTheses/a-framework-for-process-evaluations-of-psychological-intervention [accessed 2024-08-26]
60. Ismail K, Winkley K, de Zoysa N, Patel A, Heslin M, Graves H, et al. Nurse-led psychological intervention for type 2 diabetes: a cluster randomised controlled trial (Diabetes-6 study) in primary care. Br J Gen Pract 2018;68(673):e531-e540 [FREE Full text] [doi: 10.3399/bjgp18X696185] [Medline: 30012812]
61. Chasan-Taber L. Lifestyle interventions to reduce risk of diabetes among women with prior gestational diabetes mellitus. Best Pract Res Clin Obstet Gynaecol 2015;29(1):110-122 [FREE Full text] [doi: 10.1016/j.bpobgyn.2014.04.019] [Medline: 25220104]
62. Dennison RA, Ward RJ, Griffin SJ, Usher-Smith JA. Women's views on lifestyle changes to reduce the risk of developing type 2 diabetes after gestational diabetes: a systematic review, qualitative synthesis and recommendations for practice. Diabet Med 2019;36(6):702-717 [FREE Full text] [doi: 10.1111/dme.13926] [Medline: 30723968]
63. Campbell EM, Redman S, Moffitt PS, Sanson-Fisher RW. The relative effectiveness of educational and behavioral instruction programs for patients with NIDDM: a randomized trial. Diabetes Educ 1996;22(4):379-386. [doi: 10.1177/014572179602200412] [Medline: 8846745]

XSL•FO
RenderX

64. Bumbarger B, Perkins D. After randomised trials: issues related to dissemination of evidence - based interventions. J Child Serv 2008;3(2):55-64. [doi: 10.1108/17466660200800012]
65. Hawe P, Shiell A, Riley T. Complex interventions: How "out of control" can a randomised controlled trial be? BMJ 2004;328(7455):1561-1563 [FREE Full text] [doi: 10.1136/bmj.328.7455.1561] [Medline: 15217878]
66. Gabbay RA, Añel-Tiangco RM, Dellasega C, Mauger DT, Adelman A, van Horn DHA. Diabetes nurse case management and motivational interviewing for change (DYNAMIC): results of a 2-year randomized controlled pragmatic trial. J Diabetes 2013;5(3):349-357 [FREE Full text] [doi: 10.1111/1753-0407.12030] [Medline: 23368423]
67. Curry LA, Nembhard IM, Bradley EH. Qualitative and mixed methods provide unique contributions to outcomes research. Circulation 2009;119(10):1442-1452. [doi: 10.1161/CIRCULATIONAHA.107.742775] [Medline: 19289649]
68. Evans R, Scourfield J, Murphy S. Pragmatic, formative process evaluations of complex interventions and why we need more of them. J Epidemiol Community Health 2015;69(10):925-926. [doi: 10.1136/jech-2014-204806] [Medline: 25480407]

Abbreviations

DPI: diabetes prevention intervention
GDM: gestational diabetes mellitus
MI: motivational interviewing
MRC: Medical Research Council
OGTT: oral glucose tolerance test
PE: process evaluation
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
T2DM: type 2 diabetes mellitus

Edited by T de Azevedo Cardoso; submitted 24.08.23; peer-reviewed by YL Leung; comments to author 26.01.24; revised version received 06.03.24; accepted 21.06.24; published 06.02.25.

Please cite as:
Sa'id IIM, Hotung N, Benton M, Nadal IP, Baharom A, Prina M, Nisak Mohd Yusof B, Goldsmith K, Birts S, Siew Mooi C, Forbes A, Ismail K, Chew BH
Process Evaluations of Interventions for the Prevention of Type 2 Diabetes in Women With Gestational Diabetes Mellitus: Systematic Review
Interact J Med Res 2025;14:e51718
URL: https://www.i-jmr.org/2025/1/e51718
doi:10.2196/51718
PMID:

©Iklil Iman Mohd Sa'id, Natasha Hotung, Madeleine Benton, Iliatha Papachristou Nadal, Anisah Baharom, Matthew Prina, Barakatun Nisak Mohd Yusof, Kimberley Goldsmith, Samantha Birts, Ching Siew Mooi, Angus Forbes, Khalida Ismail, Boon How Chew. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 06.02.2025. 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

Evolution of Artificial Intelligence in Medical Education From 2000 to 2024: Bibliometric Analysis

Rui Li[1], BSN; Tong Wu[2], MD, PhD

[1]Emergency Department, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Obstetrics and Gynecology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Tong Wu, MD, PhD
Department of Obstetrics and Gynecology
Tongji Hospital, Tongji Medical College
Huazhong University of Science and Technology
1095 Jiefang Road
Wuhan, 430030
China
Phone: 86 13018044651
Email: tongwu66@tjh.tjmu.edu.cn

Abstract

Background: Incorporating artificial intelligence (AI) into medical education has gained significant attention for its potential to enhance teaching and learning outcomes. However, it lacks a comprehensive study depicting the academic performance and status of AI in the medical education domain.

Objective: This study aims to analyze the social patterns, productive contributors, knowledge structure, and clusters since the 21st century.

Methods: Documents were retrieved from the Web of Science Core Collection database from 2000 to 2024. VOSviewer, Incites, and Citespace were used to analyze the bibliometric metrics, which were categorized by country, institution, authors, journals, and keywords. The variables analyzed encompassed counts, citations, H-index, impact factor, and collaboration metrics.

Results: Altogether, 7534 publications were initially retrieved and 2775 were included for analysis. The annual count and citation of papers exhibited exponential trends since 2018. The United States emerged as the lead contributor due to its high productivity and recognition levels. Stanford University, Johns Hopkins University, National University of Singapore, Mayo Clinic, University of Arizona, and University of Toronto were representative institutions in their respective fields. *Cureus*, *JMIR Medical Education*, *Medical Teacher*, and *BMC Medical Education* ranked as the top four most productive journals. The resulting heat map highlighted several high-frequency keywords, including performance, education, AI, and model. The citation burst time of terms revealed that AI technologies shifted from imaging processing (2000), augmented reality (2013), and virtual reality (2016) to decision-making (2020) and model (2021). Keywords such as mortality and robotic surgery persisted into 2023, suggesting the ongoing recognition and interest in these areas.

Conclusions: This study provides valuable insights and guidance for researchers who are interested in educational technology, as well as recommendations for pioneering institutions and journal submissions. Along with the rapid growth of AI, medical education is expected to gain much more benefits.

(Interact J Med Res 2025;14:e63775) doi:10.2196/63775

KEYWORDS

artificial intelligence; medical education; bibliometric; citation trends; academic pattern; VOSviewer; Citespace; AI

Introduction

Artificial intelligence (AI) seeks to develop innovative machines that are capable of responding in ways similar to human intelligence. Its scope encompasses robotics, language recognition, image recognition, natural language processing, and expert systems. In recent years, the integration of AI into medical education (AIME) has significantly transformed traditional teaching and learning methodologies [1,2]. It can enhance clinical reasoning training, facilitate adaptive learning,

XSL•FO
RenderX

construct innovative medical teaching platforms, and facilitate critical care simulation teaching. For instance, the AI-powered simulation mannequins called SimMan 3G PLUS allows medical students to practice their clinical skills and teamwork in a fully controlled and immersive environment. Augmented reality and virtual reality technologies enable students to gain valuable hands-on experience by recreating real-life situations, without compromising patient safety [3]. In other cases, AI helps tailor educational content based on individual students' needs and abilities, and it not only fosters academic success but also cultivates critical thinking [1,4]. AI applications have become an important tool and situation for both educators and students.

As AI continues to evolve and advance, its role in promoting medical education is expected to grow. Despite many publications that have explored AI's application in medical research, they are limited to specific clinical questions in cardiology, dentistry, oncology, etc [5,6]. No bibliometric analysis has ever been conducted to assess medical education, which is essential for medical talent reserve. Therefore, a comprehensive analysis of current scientific literature is imperative for the continued advancement of AIME.

Bibliometrics uses mathematical and statistical methodologies to evaluate scholarly productivity and publication patterns within a specific discipline. We used bibliometrics to analyze the contemporary academic landscape, research priorities, and emerging trends in AIME in the 21st century. We revealed the knowledge clusters, scientific social patterns, and evolutionary nuances from an objective standpoint. This work not only simplifies navigation for researchers across disciplines but also gives valuable guidance to newcomers in the field, directing them toward promising avenues for future research.

Methods

Study Design

The research adhered to the step-by-step guidelines of bibliometric analysis and followed the reporting guideline of bibliometric reviews in biomedical literature [7].

Data Collection

The Web of Science Core Collection database was selected as the primary tool to identify relevant publications for this study. This database is well-known for its exceptional bibliometric research capabilities across over 190 subject areas and offers manual-checked literature retrieval. A 2-step process was used to construct an effective retrieval strategy. First, we extracted search terms from relevant systematic reviews and meta-analyses. Second, these search terms were meticulously reviewed and refined by authors (TW and RL) to ensure adequate coverage of all relevant research topics. The final search strategy was formulated based on this collaborative assessment: TS=("artificial intelligence" or AI or "convolutional neural network" or "recurrent neural network" or "long short-term neural memory" or "machine learning" or "genetic algorithm" or "evolutionary algorithm" or "artificial neural network" or "support vector machine" or "fuzzy logic" or "random forest" or "deep learning" or "natural language processing" or "speech recognition" or "computer vision" or "smart robot" or "video recognition" or "image recognition") and TS=(medical education or medicine education). Data collection took place on October 1, 2024, with publication years restricted from 2000 to 2024. No language restrictions were applied during the search process. Full records of retrieved papers were exported in plain text format.

Selection of Eligible Studies

Since search engines may yield results that do not fully align with the intended criteria, it is necessary to perform a manual screening process to identify the final included literature. To address this, two authors (TW and RL) independently screened the titles and abstracts of available studies based on the following exclusion criteria: (1) publications unrelated to AI. For example, some topics are associated with artificial insemination. (2) Publications unrelated to medical education, such as engineering curriculum and patient education. Notably, a majority of these studies have applied AI technologies to clinical questions and research, and only a small number pertained to medical education. Any discrepancies between the studies selected by the authors were resolved through a consensus meeting to reach a binding decision.

Data Cleaning and Preprocessing

The records from the Web of Science Core Collection database undergo rigorous selection criteria adhere to meticulous curation standards, and it partly enhances our research quality [8]. To further ensure accuracy and consistency, we performed a thorough data cleaning process. This involved the elimination of duplicate papers, which were identified through digital objective identifiers and study titles. The biblioshiny package in RStudio was used to standardize the names of authors and institutions. Any variations in author names were consolidated. As an author usually affiliates with institutions at different times, only the institution of the author at the time of publication is retained. Synonyms, aliases, and singular or plural forms, such as "AR," "augmented reality," and "augmented realities" were cleaned using a thesaurus file. The data cleaning was conducted manually by two authors (TW and RL).

Data Analysis and Visualization

To present the knowledge structure and emerging research trends, VOSviewer (version 1.6.18; Leiden University), Incites (Clarivate), InteractiVenn (Universidade de São Paulo), and Citespace (Drexel University) were used. Incites is a research evaluation tool developed by Clarivate Analytics, and the scores of citation impact, H-index, journal impact factor, international collaborations, study influence, and immediacy index were directly obtained in Incites. The citation impact is calculated by dividing the citation impact by the number of years since publication. The H-index indicates that H papers published by an author have been cited at least H times, thereby serving as a gauge of both scholarly productivity and influence. The journal impact factor was used based on the 2023 Journal Citation Reports. International collaborations were assessed to reveal the extent of interdisciplinary cooperation among coauthors hailing from various nations, highlighting the global reach and collaborative nature of the research topic. The study's influence measures the mean influence of a study within the first 5 years

XSL•FO
RenderX

of post publication, providing a view of its long-term impact. Concurrently, the immediacy index represents the average frequency with which a study is cited in the year of its publication, offering insight into the immediate reception in the academic community. The VOSviewer software was used to cluster countries, institutions, journals, and keywords. The Citespace software was used to identify keyword clusters and citation burst time. The specific parameter settings for analysis in VOSviewer and Citespace are provided in Tables 1 and 2. The latent semantic indexing and log-likelihood ratio algorithms were used for literature clustering. The InteractiVenn tool was used to identify the specific and overlapping journals in 4 categories: count, citation, top 10% papers, and cooperative work. To compare the frequencies of keywords in the AI and medical education fields, the keywords were classified based on their basic definition, and lists of professional terms or vocabulary or terminology.

Table 1. Information for clustering using the VOSviewer.

Item	Type	Algorithm	Normalization method	Minimum document or occurrence[a]
Country	Coauthorship	Majorization	Full counting	12
Organization	Coauthorship	Majorization	Full counting	15
Journal	Citation	Majorization	—[b]	12
Keyword	Co-occurrence	Majorization	Full counting	33

[a]We set no limit on the minimum citation.

[b]Not applicable.

Table 2. Information for keyword cluster using the Citespace.

Setting	Value
Time slicing	From January 2000 to October 2024, one year per slice
Text processing	Title, abstract, author keywords, and keywords plus
Node type	Keywords
Links	Strength: cosinel Scope: within slices
Selection criteria	k=7

Ethical Considerations

The institutional review board of Tongji Hospital deemed ethical approval unnecessary for this study.

Results

Baseline Characteristics

This study initially retrieved a total of 7534 publications from the search. Figure 1 provides a detailed flowchart illustrating the publication retrieval and selection process. After title and abstract screening, 2775 publications remained for further bibliometric analysis. The academic publications underwent a relatively flat growth in publication numbers from 2000 to 2017 (Figure 2A). However, the number of annual papers exhibited exponential growth since 2018, indicating the flourished development in this field. The citation counts also demonstrated a consistent rise, surpassing 10,000 citations in 2024 (Figure 2B). These documents encompassed original studies (n=1769, 63.75%), proceeding papers (n=467, 16.83%), reviews (n=237, 8.54%), and meeting abstracts (n=67, 2.41%). Proceeding papers typically contain the latest research findings and methods, as well as techniques within a scientific field, and their high percentage indicates that AIME is receiving much attention and developing rapidly. The research areas are primarily focused on Education Scientific Disciplines (n=415), Medicine General Internal (n=334), Health Care Sciences Services (n=323), Engineering Biomedical (n=264), and Surgery (n=255). It suggested that AIME is an interdisciplinary topic that needs support from specialized educators, clinical doctors, hospital administrators, and engineers.

XSL•FO
RenderX

Figure 1. Flowchart and basic information of retrieval strategy, screen process, and analysis methods.

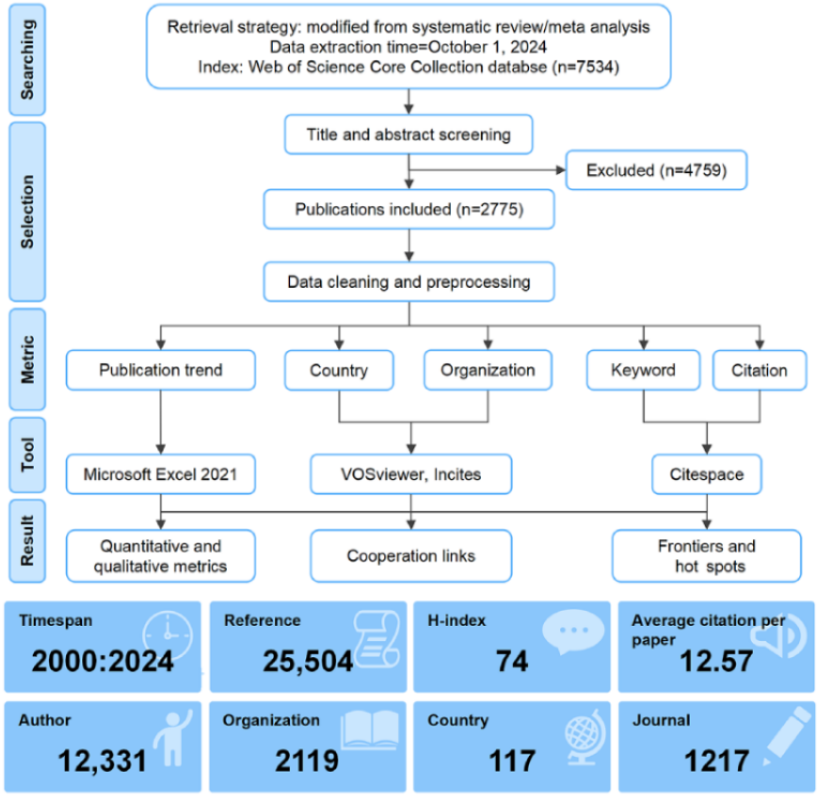
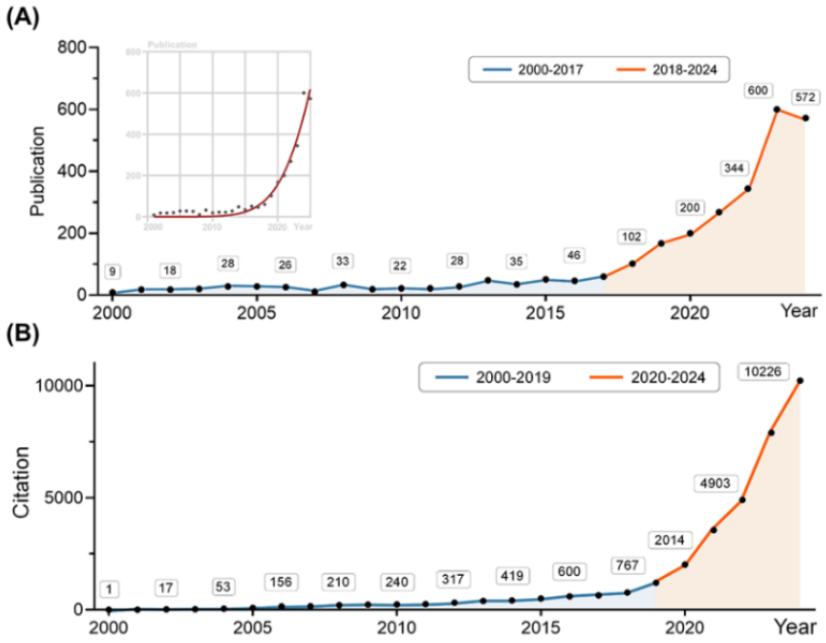


Figure 2. Baseline bibliometric characteristics of all eligible documents. (A,B) Line chart showing the annual number of documents and citations.

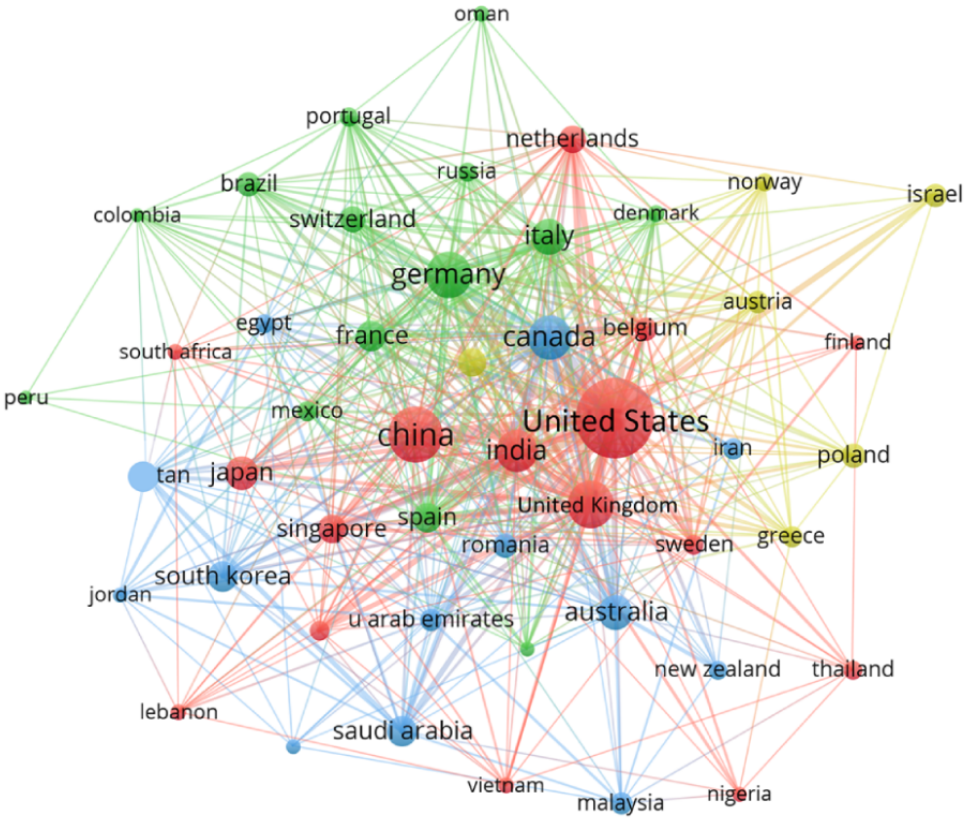


Productive and Influential Countries

In terms of national performance, 117 countries participated in the global discourse. The majority of research publications originated from Europe, North America, and Asia. The United States emerged as the primary contributor due to its high productivity (n=851 publications) and recognition levels (n=11,598 citations), exhibiting both the highest international and domestic collaborations. Although Chinese researchers published more papers than scholars from the United Kingdom (274 vs 227), their citations were relatively lower (3302 vs 3823). It may possibly be due to the limited innovation, restricted dissemination, and language barriers. The national cluster map revealed a diverse regional distribution pattern facilitated by these collaborative relationships (Figure 3). There was a strong collaboration among numerous countries, particularly close ties between the United States, China, the United Kingdom, and India.

XSL•FO
RenderX

Figure 3. Relationship and cluster of countries. Cooperative relationships networks among countries.



Institutional Performance

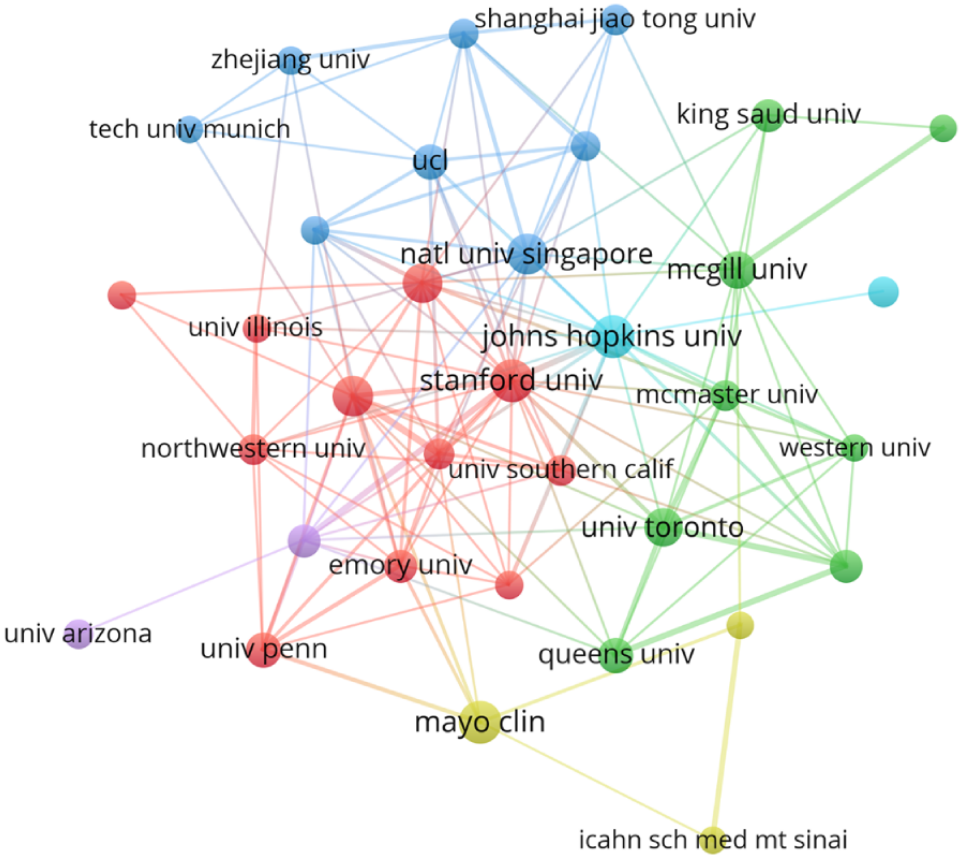
Altogether, 2119 institutions have participated in AIME studies. The top 10 institutions publishing the most papers are presented in Table 3. Among these, 4 institutions have published over 50 papers. Specifically, Harvard University emerged as the most prolific institution (n=70), and established the most international collaboration (n=35), while the University of London had the highest citation impact (n=27). The collaborative connections showed six clusters of countries publishing more than 40 papers (Figure 4). Stanford University, Johns Hopkins University, National University of Singapore, Mayo Clinic, University of Arizona, and University of Toronto were representative institutions in their respective cluster. Interestingly, regarding the impact relative to the world, the Alibaba Group had the highest score (value=61.43), followed by SUNY Downstate Health Sciences University (value=32.17), and Bukovinian State Medical University (value=25.14).

Table 3. Top 10 productive institutions of artificial intelligence in the medical education field.

Institution	Document	Citation	Citation Impact	International collaboration	H-Index	Impact relative to world
Harvard University	70	1085	15.50	35	17	1.25
University of California System	60	861	14.35	17	16	1.16
University of London	52	1211	23.29	27	17	1.88
Johns Hopkins University	51	1003	19.67	21	17	1.59
Stanford University	48	834	17.38	22	16	1.40
Harvard Medical School	47	588	12.51	25	13	1.01
Mayo Clinic	44	401	9.11	12	10	0.74
National University of Singapore	43	433	10.07	24	12	0.81
University of Toronto	39	624	16.00	14	14	1.29
University of Michigan	39	308	7.90	19	9	0.64

XSL•FO
RenderX

Figure 4. Visualization of institutional cooperation. Cooperative relationships networks.



Participants of the Source Journals

Cureus (n=78), *JMIR Medical Education* (n=72), *Medical Teacher* (n=54), and *BMC Medical Education* (n=54) ranked as the top 4 most productive journals among the 1217 involved journals (Figure 5A and Table 4). *JMIR Medical Education* published the most high-impact papers (n=48), and the second one, Cureus, only published 26. It is reasonable as *JMIR Medical Education* lies in the Q1 quartile, and likely receives more high-quality submissions, whereas Cureus is in the Q3 quartile. Cureus, *JMIR Medical Education*, *Medical Teacher*, and *International Journal of Computer Assisted Radiology and Surgery* were overlapped using a Venn diagram (Figure 5B). It was worth highlighting that Cureus and *JMIR Medical Education* had just begun to publish AIME-associated papers in 2020 (Figure 5C). Their open access characteristic helped them receive much attention in recent years. The publishers were diverse among the top 10 publication sources. More than half of these journals (7/10) published top 10% papers over 10. Notably, a significant proportion of the journals (n=1460, 52.61%) offered open access publishing, with gold access accounting for 33.26% (n=923) of this share.

XSL•FO
RenderX

Figure 5. Participated journals concerning artificial intelligence in the medical education field. (A) Relationship of representative journals. (B) Three significant journals were identified through a Venn diagram analysis of top publication volume, citation, top 10%, and cowork. (C) Annual academic output for the top five journals.

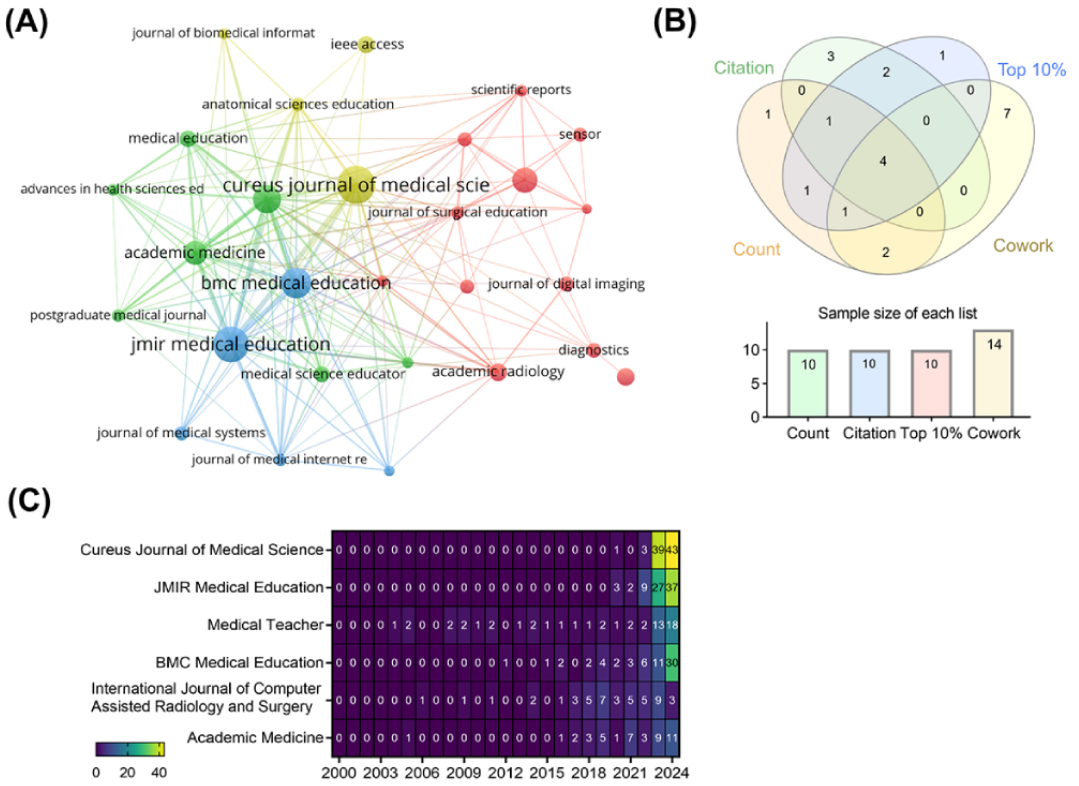


Table 4. Top five journals ranked by the counts of publications.

	Cureus	*JMIR Medical Education*	*Medical Teacher*	*BMC Medical Education*	*International Journal of Computer-Assisted Radiology And Surgery*
Document, n	78	72	54	54	46
Citation, n	910	751	1139	494	733
Top 10% paper	26	48	24	22	8
Publisher	Springer Nature	JMIR Publications Inc	Taylor & Francis Ltd	BMC	Springer Heidelberg
Cited half-life	2.5	2.1	9.1	4.4	4.8
Journal impact factor	1.1	3.2	4.7	3.4	3.1
Study influence	0.26	—[a]	1.53	0.86	0.74
Immediacy index	0.2	8.7	0.9	0.4	0.4
Open access	100%	97.22%	96.3%	24.07%	36.96%
JCI rank	173/329	18/85	10/85	15/175	67/204
JCI quartile	Q3	Q1	Q1	Q1	Q2

[a]Not applicable.

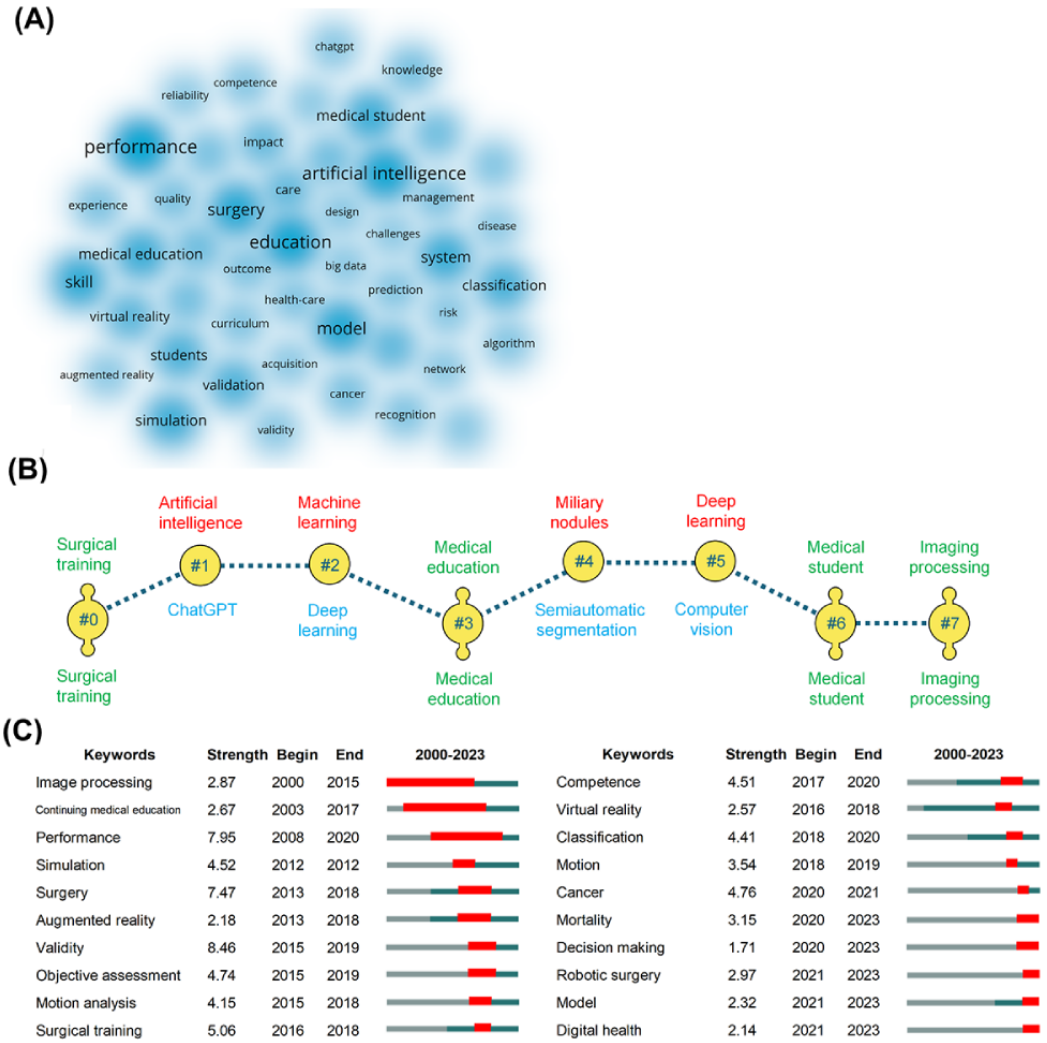
Keywords of Research Hot Spots

The study identified 7773 keywords from the titles and abstracts of the research materials, reflecting the central themes, areas of interest, and potential future developments within the discipline. The heat map highlighted high-frequency keywords, including performance (n=139), education (n=123), artificial intelligence (n=118), and model (n=103; Figure 6A). We next classified keywords according to the AI and medical education associated categories. The model (n=103), system (n=82), virtual reality (n=51), recognition (n=29), and algorithm (n=27) were popular topics, while most education focus was put on surgery (n=100), skill (n=90), simulation (n=75), classification (n=68), and diagnosis (n=47). The subsequent analysis produced eight distinct keyword clusters using log-likelihood ratio and latent semantic indexing algorithms (Figure 6B). Both algorithms

XSL•FO
RenderX

identified #0 surgical training, #3 medical education, #6 medical student, and #7 imaging processing as hub keywords. It suggested the critical role of AI applications in surgical operations and clinical image information. Citespace burst detection could reflect the research trends and innovative directions (Figure 6C). Notably, the citation burst time of terms revealed that AI technologies shifted from imaging processing (2000), augmented reality (2013), and virtual reality (2016) to decision-making (2020) and model (2021). Keywords such as mortality and robotic surgery persisted into 2023, suggesting the ongoing recognition and interest in these areas. Digital health involves the use of health apps, wearable equipment, and communication tools, which have been incorporated with AI. These devices not only provide personalized educational experiences and support mental health but also play a significant role in clinical teaching, disease prevention, and health promotion.

Figure 6. Keywords showing the hotspots and clusters. (A) Top keywords in the AIME field. (B) Eight clusters were obtained and unfolded using the latent semantic indexing (red), log-likelihood ratio (black) algorithms, or shared by both (green). (C) Keyword burstness panel showing the strength and duration. AIME: artificial intelligence into medical education.



Discussion

Principal Findings

AIME publication exhibits exponential growth these years, from approximately 50 counts in 2017 to 600 in 2023. This surge is further bolstered by the development of generative AI tools, and related national-level policies support. Here, we discuss the applications of AIME according to the keywords.

Machine Learning

Machine learning (ML), a subset of AI, is a technology that autonomously discerns inner patterns and relationships without explicit human instructions. For instance, after analyzing a large collection of cat and dog images, ML can identify distinguishing features, and subsequently differentiate between a cat and a dog in new photographs. The growing prominence of ML in medical education is evidenced by the surge in related topics, the approval of ML-based products, and the proliferation of entrepreneurial initiatives. ML-centric technologies have already been applied in clinical decision support systems, teaching tools, simulation, and training. An ML-based clinical decision support system could effectively integrate general information about HIV patients, including demographics, medical history, CD4+ lymphocyte count, viral load, genotypic data, and treatment history, to recommend an optimal combination antiretroviral

XSL•FO
RenderX

therapy [9]. More personalized advice on the appropriate dosage or duration of treatment will be provided with the help of other advanced ML algorithms [10]. In this context, it is vital to educate the next generation of medical professionals with adequate ML knowledge, enabling them to incorporate the outputs of ML tools into clinical decision-making, becoming part of this emerging data science revolution [11]. In the future, inexperienced medical students will use evidence-based learning models, like IBM's Watson Oncology system, as ordinary tools to interpret clinical data, make informed decisions, and recommend cancer treatments, in highly accordance with multidisciplinary teams [12].

Medical students require educational feedback to understand their performance and identify areas for improvement [13]. Traditional educational evaluation after surgical training falls short of providing timely, adequate, and objective assessment. A human rater is usually required to observe the video review to give a written or oral evaluation. The feedback is subjective and time-consuming, as well as limited to visual observation. With the help of automated ML-based AI assessments, the surgeon's performance can be objectively captured in a less resource-intensive way than human grading [14]. The motion, force, and vibration of robotic instruments are recorded according to the recognized structural metrics, like the Global Evaluative Assessment of Robotic Skill or the Objective Structured Assessment of Technical Skill [15].

Deep Learning

Deep learning (DL), a subset of ML, is often used to tackle complex tasks such as visual recognition, speech recognition, and natural language processing. This is achieved through the use of advanced architectures like convolutional neural networks, deep belief networks, and stacked auto-encoder networks. DL has been applied across medical undergraduate education, postgraduation education, and continuing education. A notable example is its use during medical retina rotations, where residents may not be fully trained due to limited time or access to complete their learning objectives. To address this, a DL-based model was developed to analyze a vast collection of retina images sourced from three public datasets, subsequently creating a comprehensive dataset for residents. These images were then distributed to each resident to aid in diagnosis, differential diagnosis, and therapeutic planning. The allocation system is tailored to each resident's case history, academic level, and performance, ensuring that those struggling with specific cases receive additional exposure to similar retinal conditions [16]. This approach enables AI models to identify the residents who will derive the most benefit from particular clinical cases, thereby significantly enhancing individualized ophthalmology education.

Additionally, DL technologies have been used to predict the difficulty of medical licensing examination questions, promoting more accurate assessments of examination difficulty. However, although DL models can effectively differentiate between cats and dogs, they do so by analyzing potentially hundreds or thousands of variables. The complexity and opacity of these variables often render them incomprehensible to humans. Consequently, there is an urgent need for improved interpretability methods in future DL applications to enhance understanding and transparency.

Natural Language Processing

Natural language processing plays a crucial role in smart health because of its ability to analyze and comprehend human language. The ChatGPT, developed by OpenAI Corporation, serves as a monumental tool for the natural language processing application. As for AIME, GPT-4 has been used for teaching cases, student analysis, creative writing, personalized learning guidance, and psychological support. ChatGPT is demonstrated to be effective in generating surgical procedure summaries, and its performance cannot be distinguished from a board-certified surgeon by less experienced residents [17]. In other cases, ChatGPT is used to teach the skills of breaking bad news [18], reasoning-based multiple-choice questions [19], and qualified examinations, including the Situational Judgement Test for final-year medical students in the United Kingdom [20]. These capabilities have revolutionized medical education and contributed to the overall improvement of health care delivery.

Segmentation

Image segmentation is a crucial AI technology in the radiology and pathology fields to accurately identify and delineate regions of interest, such as tumor lesions, ischemic tissues, and subcellular structures [21]. Traditional manual segmentation is not only time-consuming for physicians to learn and practice but also results in measurement variability that heavily depends on the observer's experience. In contrast, computer-assisted segmentation methods reduce the subjectivity and variability inherent in manual approaches, decrease processing time, and require minimal training. The use of image segmentation algorithms has led to significant improvements in the sensitivity and efficiency of detecting pulmonary nodules. Furthermore, these algorithms have been shown to enhance learning interest, bolster self-directed learning capabilities, sharpen problem-solving skills, and foster innovative thinking among medical trainees [22].

Effective surgical education for young surgeons presents significant challenges in practice. A segmentation-based system can identify key anatomical structures such as arteries, lymph nodes, and nerves during rectal cancer surgeries. Studies have shown the positive educational impact of AI-assisted videos in surgical training [23]. Some other AI technologies can facilitate surgical navigation and detect adverse events [24]. Taken together, real-time object segmentation is expected to play a major role in surgical education.

Comparison to Prior Work

AI has been increasingly used to enhance diagnostic accuracy, improve patient care, and facilitate the development of new treatments across various medical fields. In cardiology, for example, AI-powered algorithms are used to analyze electrocardiograms and detect arrhythmias, while in dentistry, AI-assisted diagnostic tools have been developed to identify oral pathologies [25,26]. Similarly, in oncology, AI-driven platforms are used to predict tumor growth and response to treatment [27]. Rather than concentrating on specific clinical questions, this study adopts a broader perspective to investigate

XSL•FO
RenderX

the role of AI in medical education as a whole. High-quality medical education is the key to improving the level of medical care. Through our bibliometric analysis, research trends and hot topics in this field were identified, offering insights into how AI is transforming medical education globally.

Future Direction

AIME has witnessed a remarkable evolution from the initial enthusiasm phase to the current acceptance situation [2]. Although the potential of AI technologies is expected to revolutionize medical education, related ethical considerations and challenges should be carefully examined. One major concern is whether AI might diminish the competence of medical students by increasing their reliance on external tools. Additionally, there is apprehension about the possibility of AI completely replacing medical educators. Issues such as bias, hallucinations, and uncertainties have further contributed to hesitancy in the acceptance and practical application of AI in the medical field [28]. It was necessary for a balanced approach to ensure sustainable implementation and find practical ways to incorporate AI into curricula. As deeper investigation is conducted, AI will be an integral part of medical education, highlighting a journey of personal and professional growth alongside technological adoption.

Since the release of ChatGPT, AI-generated content (AIGC) has emerged as an innovative educational tool with significant potential. AIGC technologies have the capability to reshape pedagogical practices through various applications, such as virtual patient construction, automated question bank generation, and 3D anatomical simulations [19,29]. These technologies may also help address the uneven distribution of educational resources and facilitate the updating of outdated knowledge. Exploring the application of AIGC in medical education represents both a transformation and expansion of existing teaching models, as well as a forward-looking exploration of future cultivation patterns.

AIME introduces new demands for educational accreditation systems and training models. While traditional medical education accreditation emphasizes faculty strength, physical facilities, and curriculum design, AI enriches these standards by incorporating digital classes, digital resources, and dynamic monitoring. In the context of postgraduate education, AIME necessitates a re-evaluation of training duration, content, and assessment methods. It is crucial for policy makers and practitioners to explore how AI can be reasonably integrated into residents' learning, ensuring a balance that prevents over-reliance on technology and maintains the importance of clinical practical experience.

Limitations

This study has several limitations. First, the bibliometric analysis was conducted exclusively using the Web of Science Core Collection database. This singular focus may introduce bias into the results, as it does not account for data from other significant sources. Future research should incorporate additional databases, such as Scopus and Google Scholar, to provide a more comprehensive analysis. Second, while citation frequency is often used as a measure of academic influence, it does not necessarily equate to positive evaluations. Citations can also be made for criticism or rebuttal. Therefore, supplementing bibliometric analysis with content analysis would offer a more accurate reflection of the literature's true impact. Finally, bibliometric analysis inherently concentrates on published research, which may not capture the most current research trends and developments.

Conclusions

This study delves into the current landscape of AI applications in medical education, encompassing the geographical distribution of research efforts, recognition of pivotal researchers, identification of key research trends, and exploration of emerging domains. There is a burgeoning interest in AIME and an expanding comprehension of its potential impact. The United States has emerged as a leader in this field, with many institutions standing out as prolific organizations. As the demand for more personalized and effective medical education grows, there is a pressing need for large-scale, rigorously designed studies to provide empirical evidence of AI's effectiveness and safety.

Acknowledgments

This work is financially supported by grants from the Teaching and Research Project of the Second Clinical College of Tongji Medical College (TJXYJ2023016 and TJSZ2024016). During the preparation of this work, the authors used Acadwrite [30] to improve readability, grammar, and language style. After using this tool, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

Data Availability

All data generated or analyzed during this study are included in this published article.

Authors' Contributions

RL designed the study and wrote the initial draft of the manuscript. TW supervised the study, contributed to manuscript review and editing, and provided the funding. All authors contributed to the methodology and data visualization. All authors have read and agreed to the published version of the manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Cooper A, Rodman A. AI and medical education—a 21st-century pandora's box. N Engl J Med 2023;389(5):385-387. [doi: 10.1056/NEJMp2304993] [Medline: 37522417]
2. Gordon M, Daniel M, Ajiboye A, Uraiby H, Xu NY, Bartlett R, et al. A scoping review of artificial intelligence in medical education: BEME guide No. 84. Med Teach 2024;46(4):446-470 [FREE Full text] [doi: 10.1080/0142159X.2024.2314198] [Medline: 38423127]
3. Montoya-Rodríguez MM, de Souza Franco V, Llerena CT, Cobos FJM, Pizzarossa S, García AC, et al. Virtual reality and augmented reality as strategies for teaching social skills to individuals with intellectual disability: a systematic review. J Intellect Disabil 2023;27(4):1062-1084. [doi: 10.1177/17446295221089147] [Medline: 35420906]
4. Bhatnagar A, Kekatpure AL, Velagala VR, Kekatpure A. A review on the use of artificial intelligence in fracture detection. Cureus 2024;16(4):e58364 [FREE Full text] [doi: 10.7759/cureus.58364] [Medline: 38756254]
5. Lu W, Yu X, Li Y, Cao Y, Chen Y, Hua F. Artificial intelligence-related dental research: bibliometric and altmetric analysis. Int Dent J 2024 Sep 11 (forthcoming) [FREE Full text] [doi: 10.1016/j.identj.2024.08.004] [Medline: 39266401]
6. Suero-Abreu GA, Hamid A, Akbilgic O, Brown SA. Trends in cardiology and oncology artificial intelligence publications. Am Heart J Plus 2022;17:100162 [FREE Full text] [doi: 10.1016/j.ahjo.2022.100162] [Medline: 38559882]
7. Montazeri A, Mohammadi S, Hesari PM, Ghaemi M, Riazi H, Sheikhi-Mobarakeh Z. Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): a minimum requirements. Syst Rev 2023;12(1):239 [FREE Full text] [doi: 10.1186/s13643-023-02410-2] [Medline: 38102710]
8. Lim WM, Kumar S, Donthu N. How to combine and clean bibliometric data and use bibliometric tools synergistically: guidelines using metaverse research. J Bus Res 2024 Sep;182:114760. [doi: 10.1016/j.jbusres.2024.114760]
9. Parbhoo S, Bogojeska J, Zazzi M, Roth V, Doshi-Velez F. Combining kernel and model based learning for HIV therapy selection. AMIA Jt Summits Transl Sci Proc 2017;2017:239-248 [FREE Full text] [Medline: 28815137]
10. Peiffer-Smadja N, Rawson T, Ahmad R, Buchard A, Georgiou P, Lescure FX, et al. Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clin Microbiol Infect 2020;26(5):584-595 [FREE Full text] [doi: 10.1016/j.cmi.2019.09.009] [Medline: 31539636]
11. James C, Wheelock KM, Woolliscroft JO. Machine learning: the next paradigm shift in medical education. Acad Med 2021;96(7):954-957. [doi: 10.1097/ACM.0000000000003943] [Medline: 33496428]
12. Jie Z, Zhiying Z, Li L. A meta-analysis of watson for oncology in clinical application. Sci Rep 2021;11(1):5792 [FREE Full text] [doi: 10.1038/s41598-021-84973-5] [Medline: 33707577]
13. González A. Turning a traditional teaching setting into a feedback-rich environment. Int J Educ Technol High Educ 2018 Aug 27;15(1):32. [doi: 10.1186/s41239-018-0114-1]
14. Brown JD, Kuchenbecker KJ. Effects of automated skill assessment on robotic surgery training. Int J Med Robot 2023 Apr;19(2):e2492. [doi: 10.1002/rcs.2492] [Medline: 36524325]
15. Nielsen PE, Foglia LM, Mandel LS, Chow GE. Objective structured assessment of technical skills for episiotomy repair. Am J Obstet Gynecol 2003;189(5):1257-1260. [doi: 10.1067/s0002-9378(03)00812-3] [Medline: 14634550]
16. Muntean GA, Groza A, Marginean A, Slavescu RR, Steiu MG, Muntean V, et al. Artificial intelligence for personalised ophthalmology residency training. J Clin Med 2023;12(5):1825 [FREE Full text] [doi: 10.3390/jcm12051825] [Medline: 36902612]
17. Lebhar MS, Velazquez A, Goza S, Hoppe IC. Dr. ChatGPT: utilizing artificial intelligence in surgical education. Cleft Palate Craniofac J 2024;61(12):2067-2073. [doi: 10.1177/10556656231193966] [Medline: 37545428]
18. Webb J. Proof of concept: using ChatGPT to teach emergency physicians how to break bad news. Cureus 2023;15(5):e38755 [FREE Full text] [doi: 10.7759/cureus.38755] [Medline: 37303324]
19. Agarwal M, Sharma P, Goswami A. Analysing the applicability of ChatGPT, Bard, and Bing to generate reasoning-based multiple-choice questions in medical physiology. Cureus 2023;15(6):e40977 [FREE Full text] [doi: 10.7759/cureus.40977] [Medline: 37519497]
20. Borchert RJ, Hickman CR, Pepys J, Sadler TJ. Performance of ChatGPT on the situational judgement test—a professional dilemmas-based examination for doctors in the United Kingdom. JMIR Med Educ 2023;9:e48978 [FREE Full text] [doi: 10.2196/48978] [Medline: 37548997]
21. Glotsos D, Spyridonos P, Cavouras D, Ravazoula P, Dadioti P, Nikiforidis G. Automated segmentation of routinely hematoxylin-eosin-stained microscopic images by combining support vector machine clustering and active contour models. Anal Quant Cytol Histol 2004;26(6):331-340. [Medline: 15678615]
22. Duong MT, Rauschecker AM, Rudie JD, Chen PH, Cook TS, Bryan RN, et al. Artificial intelligence for precision education in radiology. Br J Radiol 2019;92(1103):20190389 [FREE Full text] [doi: 10.1259/bjr.20190389] [Medline: 31322909]

XSL•FO
RenderX

23. Kinoshita K, Maruyama T, Kobayashi N, Imanishi S, Maruyama M, Ohira G, et al. An artificial intelligence-based nerve recognition model is useful as surgical support technology and as an educational tool in laparoscopic and robot-assisted rectal cancer surgery. Surg Endosc 2024;38(9):5394-5404. [doi: 10.1007/s00464-024-10939-z] [Medline: 39073558]
24. Bamba Y, Ogawa S, Itabashi M, Shindo H, Kameoka S, Okamoto T, et al. Object and anatomical feature recognition in surgical video images based on a convolutional neural network. Int J Comput Assist Radiol Surg 2021;16(11):2045-2054 [FREE Full text] [doi: 10.1007/s11548-021-02434-w] [Medline: 34169465]
25. Aldakhil S, Alkhurayji K, Albarrak S, Almihbash A, Aldalan R, Alshehri K, et al. Awareness and approaches regarding artificial intelligence in dentistry: a scoping review. Cureus 2024;16(1):e51825 [FREE Full text] [doi: 10.7759/cureus.51825] [Medline: 38327934]
26. Siontis KC, Noseworthy PA, Attia ZI, Friedman PA. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management. Nat Rev Cardiol 2021;18(7):465-478 [FREE Full text] [doi: 10.1038/s41569-020-00503-2] [Medline: 33526938]
27. Hoang DT, Dinstag G, Shulman ED, Hermida LC, Ben-Zvi DS, Elis E, et al. A deep-learning framework to predict cancer treatment response from histopathology images through imputed transcriptomics. Nat Cancer 2024;5(9):1305-1317. [doi: 10.1038/s43018-024-00793-2] [Medline: 38961276]
28. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. N Engl J Med 2018;378(11):981-983 [FREE Full text] [doi: 10.1056/NEJMp1714229] [Medline: 29539284]
29. Stokel-Walker C, Van Noorden R. What ChatGPT and generative AI mean for science. Nature 2023;614(7947):214-216. [doi: 10.1038/d41586-023-00340-6] [Medline: 36747115]
30. Acadwrite. URL: https://acadwrite.cn/ [accessed 2025-01-16]

Abbreviations

AI: artificial intelligence
AIGC: artificial intelligence–generated content
AIME: artificial intelligence into medical education
DL: deep learning
ML: machine learning

Edited by T de Azevedo Cardoso; submitted 29.06.24; peer-reviewed by W Yang, W Qi, Y Jiang; comments to author 19.09.24; revised version received 04.11.24; accepted 26.11.24; published 30.01.25.

Please cite as:
Li R, Wu T
Evolution of Artificial Intelligence in Medical Education From 2000 to 2024: Bibliometric Analysis
Interact J Med Res 2025;14:e63775
URL: https://www.i-jmr.org/2025/1/e63775
doi:10.2196/63775
PMID:

©Rui Li, Tong Wu. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 30.01.2025. 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 Recruitment of Adolescents Aged 16-18 Years in a Web-Based Peer Network Study Through Financial Reimbursements: Randomized Controlled Trial

Sarah Eddy[1]; Rachel Sacks-Davis[1,2,3]; Michelle Raggatt[1,2]; Cassandra Wright[1,4]; Paul Dietze[1,2,5]; Margaret Hellard[1,2,6]; Jane Hocking[3]; Megan S C Lim[1,2,3]

[1]Burnet Institute, 85 Commercial Rd, Melbourne, Australia

[2]School of Population Health and Preventive Medicine, Monash University, Melbourne, Australia

[3]School of Population and Global Health, University of Melbourne, Melbourne, Australia

[4]Menzies School of Health Research, Darwin, Australia

[5]National Drug Research Institute, Curtin University, Perth, Australia

[6]Alfred Health, Melbourne, Australia

Corresponding Author:
Megan S C Lim
Burnet Institute, 85 Commercial Rd, Melbourne, Australia

Abstract

Background: Peers are known to influence the health behaviors and attitudes of adolescents, yet recruitment of these networks is challenging. Previous studies have used web-based respondent-driven sampling methods to recruit this population, yet none have experimentally investigated the impact of financial reimbursements.

Objective: This study aimed to (1) compare the effectiveness of two financial reimbursement strategies for recruiting adolescents and their peer networks and (2) explore factors associated with successfully recruiting peers.

Methods: A parallel-design randomized controlled trial was conducted in which participants (seeds) were randomly allocated to a fixed cash reimbursement (control) or scaled reimbursement (experimental) group as a strategy to be recruited into a web-based peer network study. Seeds aged 16 to 18 years were recruited through social media advertisements and an online student panel. They completed a web-based survey, which assessed eligibility and included questions about their friends (peers). Allocation occurred through a survey platform using a simple randomization method. In the fixed group, all participants in a peer network received AUD $5 (US $3.29); in the scaled group, all participants in a peer network received an additional AUD $5 (US $3.29) per peer who successfully completed the survey (up to AUD $30 each [US $19.72]). Participants and researchers were not blinded to intervention groups. The primary outcome was recruitment of peers to complete the web-based survey (proportion of nominated peers). The number of peers recruited was a secondary outcome. In secondary analyses, we identified peer-, relationship-, and seed-level variables associated with successfully recruiting peers.

Results: Of 463 seeds allocated to an intervention (scaled n=221 and fixed n=242), 319 (68.9%) had complete data for analysis (scaled n=157, 71% and fixed n=162, 67%). A total of 11.9% of seeds successfully referred peers (18.5% scaled group and 5.6% fixed group). Those in the scaled reimbursement intervention were 3.80 times more likely to successfully recruit their peers than those in the fixed reimbursement intervention (proportion ratio 3.80, 95% CI 1.78 - 8.09). Similarly, the average number of peers recruited differed by 0.19 (95% CI 0.11 - 0.28) per seed between the scaled and fixed intervention groups. Peer recruitment success was similar regardless of the gender, age, education level, and network size of seeds or the gender, age, and closeness of peers. Seeds recruited through social media were more likely to successfully recruit their nominated peers than those recruited through a research panel (proportion ratio 2.20, 95% CI 1.06 - 4.55).

Conclusions: Scaled reimbursements resulted in significantly greater recruitment of peers than fixed reimbursements; however, the total number of peers recruited was low. Greater-value incentives and stronger initial recruitment through social media may be needed to recruit large numbers of friend networks.

International Registered Report Identifier (IRRID): RR2-10.2196/44813

(Interact J Med Res 2025;14:e70223) doi:10.2196/70223

KEYWORDS

financial reimbursement; incentive; adolescent; peer; young person; web-based respondent-driven sampling; peer network; social network

XSL•FO
RenderX

Introduction

Background

The period from adolescence to adulthood is a significant period of transition during which key health risk behaviors emerge. This time is characterized by experimentation with alcohol and other drugs and sexual behaviors and the onset of mental health issues [1]. Researchers have consistently highlighted the importance of the social context in understanding the development of both positive and negative adolescent health behaviors [2,3]. As emphasized in social network theory, an individual's social network influences their attitudes and behaviors [3]. Peer influence is thought to occur through different mechanisms, including homophilic social selection—the selection of friends based on shared behaviors or attitudes—and social influence—the convergence of similar behaviors and attitudes within a network over time [4]. The presence of peers can increase adolescent susceptibility to risky behaviors, including sexual risk taking and alcohol use [5]. There is extensive evidence of high similarity in health behaviors among networks of young people [4,6-8]. By studying adolescents in the context of their peer networks, health interventions can be designed that utilize these social networks.

To date, studies investigating adolescent social networks have examined health behaviors such as smoking, alcohol and substance use, physical activity, and dieting [4]. However, these studies have predominantly taken place in school settings, using an existing dataset (1994 - 2002) from the United States [4]. While it is convenient and useful to sample adolescents and their friendships within schools, this approach does not reach those disengaged from mainstream education or capture the transitory period after completing school. Less is known about the influence of nonschool peer relationships, particularly relevant in the current social media era where relationship boundaries are less defined. On social media, data are inherently organized in a network and provide an opportunity to study these processes at a larger scale [9].

Respondent-driven sampling (RDS) is a recruitment method typically used for hard-to-reach populations, where *seeds* (primary participants) are recruited and then asked to recruit several *peers* (secondary participants) within their network [10]. In recent years, there has been an increase in the use of web-based RDS (WebRDS), an online version of RDS that allows participants to be recruited without the need for face-to-face contact [11]. This method confers significant advantages over traditional RDS, including ease of access, reductions in time and cost, and respondent privacy and anonymity [10,12]. Equally, researchers have identified drawbacks of the web-based approach, including the increased potential for fraudulent and duplicate responses, lower quality data, and key differences in the population recruited (eg, higher socioeconomic status, fewer Aboriginal and Torres Strait Islander participants, and more participants who speak a language other than English at home) [10,12]. Despite these challenges, the nature of WebRDS provides an opportunity to study the interactions between and within networks in an online format. Existing studies have recruited seeds through social media ads (eg, Facebook), through online research panels, or through the researcher's networks and often include some researcher contact (ie, via phone or email) to confirm eligibility or motivation to engage [11]. Seeds typically recruit their peers by sharing a link via direct messaging (Facebook or WhatsApp) or email (see Helms et al [11] for a review). Since young people's lives largely exist online, and those engaging in higher risk behaviors are difficult to reach through traditional recruitment methods, RDS is well suited to recruiting young people and their peer networks.

While recruitment of peer networks can be challenging, one factor typically associated with successful recruitment is financial reimbursement, whereby seeds receive additional reimbursement for each peer they recruit [11]. In WebRDS studies conducted with adolescent and youth samples, seed reimbursement has ranged from US $4 to 30, with an additional incentive of US $4 to 20 per peer successfully recruited (maximum 2-5 referrals) [12-16]. Many studies have shown that offering incentives significantly improves individual participation in research [17]. However, to our knowledge, no studies to date have compared the effectiveness of different incentives for recruiting peers among adolescents and young people. One study investigated the best way to use internet-based RDS to recruit friends into an existing research panel within the general population and found that the combination of higher incentives for each peer recruited and the inclusion of a sign-up bonus (for joining the research panel) was effective at increasing the number of friends recruited [18]. The preference for incremental incentives (ie, an additional reimbursement for each peer) over a fixed amount (regardless of the number of peers recruited) was also emphasized in an adult sample of men who have sex with men [19]. Notably, this study reported that monetary incentives were more important for younger participants [19]. It remains unclear how varying financial reimbursements affect a young person's willingness to refer peers.

Demographic, relational, and methodological factors may also have an impact on recruitment success. While often seeds are more likely to recruit peers with similar demographic characteristics due to homophilic selection, this does not always explain recruitment differences [20]. Some studies within the general population have shown that productive seeds (in terms of recruitment or referrals) are more likely to be younger, female, more highly educated, and with a larger peer network [21]. Online recruitment and a smaller number of initial seeds have also been identified as key factors [11]. Further research within a youth sample is needed to understand the characteristics of successful seeds and their referred peers.

Rationale

We conducted a randomized controlled trial that aimed to determine the effectiveness of using WebRDS to recruit adolescent peer networks. Importantly, this study addresses a gap in the literature by investigating the impact of two different reimbursement structures on recruitment success—a scaled reimbursement (experimental) group, in which all participants in a peer network received an additional AUD $5 (US $3.29) for each peer who was successfully recruited, to a maximum

XSL•FO
RenderX

value of AUD $30 each (US $19.72); and a fixed reimbursement (control) group, in which all participants were reimbursed AUD $5 (US $3.29) for completing the survey. In the context of ongoing challenges with recruiting peer networks, this study provides valuable information to determine the feasibility and cost of monetary incentives in future WebRDS studies.

Aims and Research Questions

The primary aim was to compare the effectiveness of two financial reimbursement strategies for recruiting peer networks, explored through the following research questions:

1. Does a scaled group financial reimbursement for peer referral increase the proportion of an individual's peer network who complete the survey compared to having no peer referral reimbursement?
2. Does a scaled group financial reimbursement for peer referral increase the number of peers per participant who complete the survey compared to having no peer referral reimbursement?

We hypothesized that a scaled group financial reimbursement for peer referral would result in a significantly higher proportion of participants' peer networks and a significantly higher number of peers per participant who complete the survey, compared to a fixed reimbursement rate for individuals.

A secondary aim was to determine what demographic, relational, and methodological factors were associated with successfully recruiting peers. As this was an exploratory analysis, no hypotheses were specified.

Methods

Study Design

To test the effectiveness of reimbursements for peer referral in a web-based cross-sectional peer network study, a 2-group parallel randomized controlled trial (RCT) design was used. Participants were then randomly allocated in a 1:1 ratio to the scaled group reimbursement (experimental) or the fixed group reimbursement (control). The study design was informed by qualitative interviews with young people [22]. This RCT was not registered as the primary outcome (recruitment of friends) was not a direct health or medical outcome. The trial is reported in accordance with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines for web-based RCTs [23]. Further details can be found in the published protocol [22]. A deviation from this published protocol occurred due to lack of success recruiting participants via social media. Recruitment was slower than anticipated, and a greater number of participants were recruited from Student Edge instead of social media (see Recruitment section).

Participants

In this study, seeds were defined as individuals recruited directly through online recruitment platforms who had the opportunity to recruit their friends. Peers were defined as friends nominated by seeds to complete the survey. Seeds were eligible if they were aged between 16 and 18 years and currently living in Australia; peers were eligible if they were aged 16 years and older and currently living in Australia. A peer network encompassed a seed and their peers.

Recruitment

Recruitment initially targeted participants through three platforms: Instagram or Facebook, Snapchat, and Student Edge panel (detailed further in protocol [22]). Instagram and Facebook were treated as a single recruitment source, given that they are managed through the same advertising account and owned by the parent company Meta. Student Edge is a web-based organization that hosts online surveys targeted at young people in high school and technical and further education in Australia [24]. The survey was made available to Student Edge members on their website, while recruitment through social media comprised paid advertisements. Advertisements in still and video format were designed to appeal to young people, and they only briefly mentioned the study topic and the reimbursement. Recruitment initially took place from March 21 to April 2, 2022. Initial recruitment through social media platforms posed challenges, with no participants successfully recruited through Snapchat and only small numbers recruited through Instagram or Facebook. As such, in early 2023, the advertisements were revised into video format with brighter colors and more dynamic text and hosted on TikTok as well as Instagram or Facebook, with the goal of increasing recruitment through social media platforms. TikTok is a free hosting platform for short-form videos, primarily accessed via smartphone. As was the case with Snapchat, TikTok advertising proved unsuccessful, while a small additional sample of young people was recruited through Instagram or Facebook with the revised advertisements. Following this, the remainder of the target sample was recruited through the Student Edge platform. Student Edge participants who completed the survey in 2022 were unable to complete it a second time. This second phase of recruitment occurred between March 15 and May 12, 2023. In total, across both years, active recruitment took 10 days more than anticipated.

Intervention

First, seed participants answered questions about their own behaviors and then were asked to specify their number of "close friends" (peers), up to a maximum of 10. Seed participants were then asked to answer questions about each peer, including initials and a chosen codename, and advised of the reimbursement amount depending on their reimbursement group. At the completion of the survey, seeds had the opportunity to either personalize a message or use a suggested message inviting their peers to participate in the survey. Seeds received this information in an email immediately after completing the survey and again after 72 hours. Seed participants were provided with a unique link via email to share with their peers, which included their record ID. Peers were automatically recruited in the same reimbursement group as their seed. When commencing the survey, peers were asked to provide their given codename and their initials. Peer networks were linked using the seed's ID number and peer initials and codenames. Linkage required a match of seed ID and at least one of the peer's codename or initials. Peers were also asked questions about their own peer network, yet were not able to refer them to complete the survey.

XSL•FO
RenderX

All seeds and peers in the fixed reimbursement group received AUD $5 (US $3.29) for completing the survey, with no additional reimbursement for peers in their network who successfully completed the survey. In the scaled reimbursement group, seeds and peers within a peer network received AUD $5 (US $3.29) for completing the survey, plus an additional AUD $5 (US $3.29) for each peer who successfully completed the survey, up to a maximum of AUD $30 (US $19.72) per participant (detailed in [22]). At the completion of the recruitment phase, participants were reimbursed with online gift vouchers sent via SMS text messaging through a third-party service. Due to the existing platform features, participants recruited through Student Edge received the first AUD $5 (US $3.29) through the Student Edge platform and any additional value in the form of gift vouchers. Participants could contact the research team via an online form if they encountered any issues with the survey.

Outcomes

The primary outcome was defined as completion of the survey by a peer (yes/no). We estimated the difference in the proportion of seeds' nominated peers who completed the survey in the scaled group reimbursement compared to the fixed individual reimbursement group. Survey completion was recorded if the peer clicked the final survey submission button. This outcome measure indicates the level of complete data about each seed's peer network.

The secondary outcome was the number of a seed's peers who completed the survey. We estimated the difference in the number of peers who completed the survey per seed in the scaled group reimbursement compared to the fixed individual reimbursement group. This outcome measure assesses whether the reimbursement model results in an overall higher number of participants.

Randomization

Randomization (generation and allocation) occurred automatically through the REDCap survey platform [25]. This involved a simple randomization based on the final digit (seconds) of the main survey page timestamp. Those who commenced the survey with an even number of seconds were allocated to the scaled (experimental) group, and those with an odd number of seconds were allocated to the fixed (control) group. The intervention was identical, differing only in the amount of reimbursement offered. Randomization occurred after consent, yet participants were only advised of which intervention group they had been allocated to at the end of the main survey section, prior to answering questions about their peers. Researchers were not blinded to allocation on the survey platform yet did not interact with participants directly. Participants were unaware of the other experimental condition.

Data Collection

Participants completed an online survey through REDCap. Participants could not proceed with the survey if they did not meet the age criteria. The survey captured potential factors associated with successfully recruiting peers. This included seed-level factors such as demographic characteristics, recruitment setting, and network size. Peer-level variables included perceived closeness, age, and gender. To increase engagement, seeds could choose a codename and avatar to identify themselves and a codename and avatar to identify each peer. The survey also covered topics such as social media and media related to sexual health (ie, pornography, sexting, and dating apps), mental health, sexual health, alcohol and other drug use, and COVID-19 vaccination, which will be reported on elsewhere.

Sample Size and Statistical Analysis

We determined the required sample size of seeds to detect a difference in the proportion of an individual's peer network that completed the survey between each intervention group. In line with previous research, we predicted that seeds would refer an average of three peers [26,27] and that approximately half of participants in the scaled group would successfully recruit at least one peer (as indicated by the peer's completion of the survey) [12,15]. Assuming the probability of recruiting each peer is binomial and the probability of recruiting no peers being $(1-p)^n$ (where n is the total number of peers and p is the probability of any one nomination being recruited), the probability of any nomination completing the survey was 0.2. We assumed some variation between friendship groups, including an intracluster correlation of 0.5 and a mean cluster size of 3, with a coefficient of variation for cluster size of 0.83. On this basis, 153 seeds were required per intervention group (total of 306 participants). This estimate allowed us to detect a difference of 1.4 peers per seed completing the survey between intervention groups, with 80% power. Conservatively, we estimated a total of 92 peers across both groups. We intended to recruit equal numbers from each of the three recruitment settings [22], yet due to the limited success with social media advertising outlined above, a greater number of participants were recruited through the Student Edge platform.

To assess balance in demographic characteristics between seeds in the two intervention groups, descriptive statistics were reported by intervention group in table form, reporting mean and SD for continuous data and number of participants (n) and percentages for categorical data.

The primary outcome was evaluated in a modified intention-to-treat analysis where participants were excluded from analysis if they did not answer questions about their peers, as this was considered an incomplete response. The exception to this was if participants were identified as having zero close friends. To assess the effect of the reimbursement intervention on the proportion of a seed's peers who completed the survey, we used generalized estimating equations (GEE) with a log link, binomial family, and exchangeable correlation structure to account for within friendship group correlation. In this analysis, peer nominees were the unit of analysis, grouped by their seed. Each peer could either be successfully recruited or not recruited (binary outcome). This analyzed seeds who nominated at least one peer.

To determine the effect of the intervention on the number of peers per participant who completed the survey (secondary outcome), we estimated the expected difference in counts between groups using predictions from a Poisson regression

XSL•FO
RenderX

model. Poisson regression was chosen for analysis based on the distribution of the number of peers. In secondary analyses, we identified factors associated with successfully recruiting each peer, considering peer-, relationship-, and seed-level variables. Predictor selection was informed by prior studies investigating online survey recruitment [11,20,21]. We used bivariable log binomial GEE, where the outcome was the successful recruitment of a nominated peer. Potential predictors included recruitment setting, demographics, perceived closeness, and network size. Exchangeable correlation structures were used to account for clustering. In addition, we identified factors associated with recruiting a larger number of peers. We used bivariable Poisson regression models. Only seed-level variables were considered for the latter analysis. For simplicity in interpretation, demographic variables in secondary analyses were included in categorical form or converted into categories, and "I don't wish to say" responses were coded as missing.

Ethical Considerations

This study received approval from the Monash University Human Research Ethics Committee (project 23132). Participants were provided with information about the study in both video and written form. Informed consent was obtained through a series of checkboxes, indicating that participants had understood the information provided and were happy to take part and share their information for the study purpose. All data were deidentified for analysis. As described in detail earlier in the Intervention section, participants received between AUD $5 (US $3.29) and AUD $30 (US $19.72) in compensation. Details of relevant support services were provided at the completion of the survey.

Results

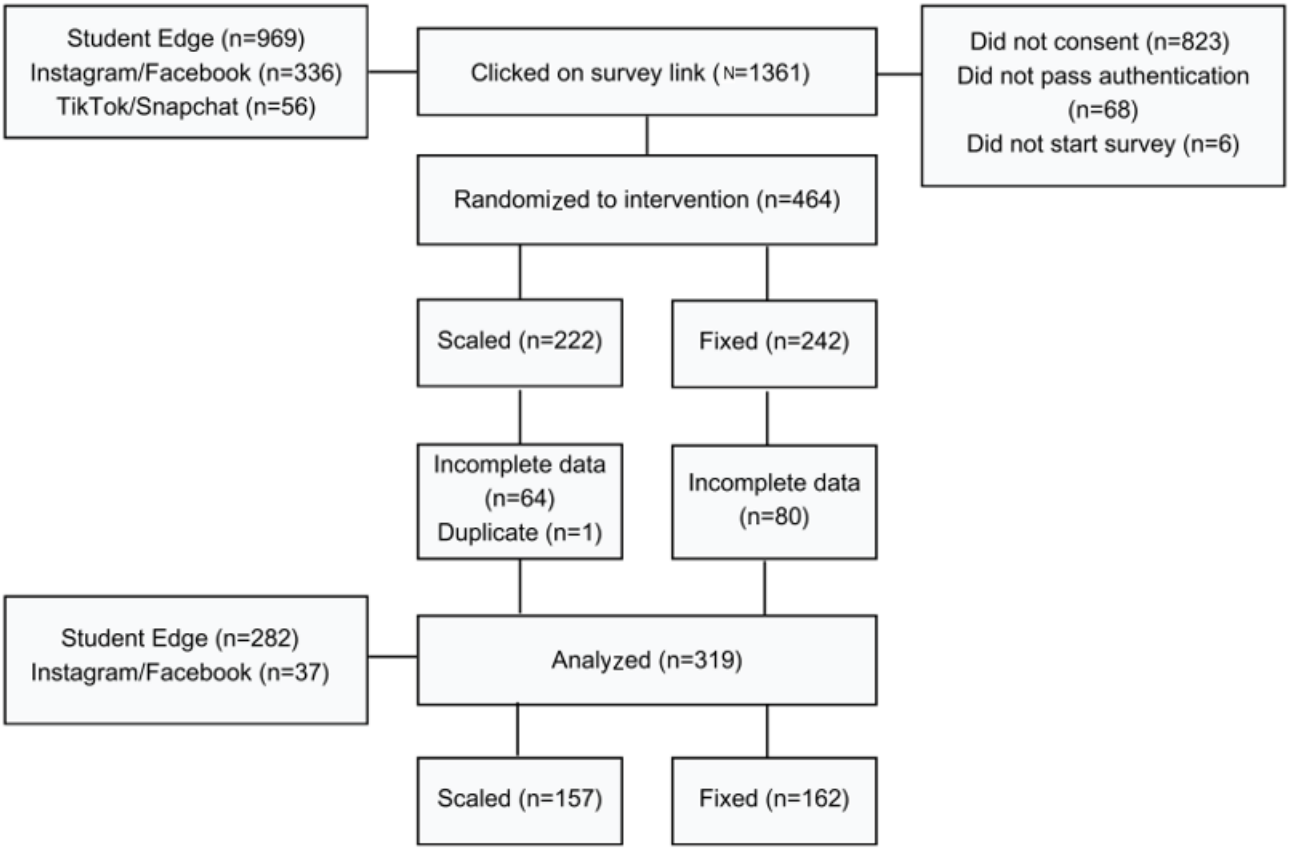
Overview

A total of 1361 seeds clicked on the survey link. Of these, a total of 897 did not provide consent, pass authentication (due to being identified as a duplicate), or commence the main survey following authentication (see Figure 1). The remaining 464 participants were randomized to the scaled intervention (n=222) or the fixed intervention (n=242). Of these, 144 only partially completed the survey, and 1 was identified as a duplicate. Thus, 319 participants were analyzed (n=157 in the scaled group and n=162 in the fixed group). Of those with partial completion, 127 stopped the survey post unblinding. Post-unblinding attrition was higher in the fixed group (n=74, 30.6%) than in the scaled group (n=53, 23.9%).

The demographic characteristics of seeds overall and separated by intervention group are shown in Table 1. Intervention groups were balanced in demographic characteristics. Overall, the sample had a mean age of 17.03 (SD 0.8) years, and 60.8% (n=194) identified as a woman. Most were from New South Wales (n=128, 40.1%) or Victoria (n=71, 22.3%), were currently studying (n=299, 66.2%), and spoke English at home (n=194, 60.8%). Most of the sample (n=282, 88.4%) were recruited through the Student Edge panel, and almost two-thirds (n=198, 62.1%) were recruited during the second wave of data collection in 2023, yet these characteristics were consistent across intervention groups. There were no missing data as participants were required to answer each question, with the option to select "I don't wish to say." Of variables included in secondary analyses, seed-level gender (n=4, 1.3%) and education level (n=9, 2.8%), and peer-level gender (n=7, 2.3%) and closeness (n=1, 0.5%), contained a small number of "I don't wish to say" responses, which were not included in the analysis.

XSL•FO
RenderX

Figure 1. Flowchart showing seed involvement in the intervention from recruitment through to analysis.



XSL•FO
RenderX

Table . Demographic characteristics of primary participants.

Variables	Scaled (n=157)	Fixed (n=162)	Total (n=319)
Age (years), mean (SD)	17.03 (0.8)	17.04 (0.8)	17.03 (0.8)
Gender, n (%)			
Man	49 (31.2)	54 (33.3)	103 (32.3)
Woman	99 (63.1)	95 (58.7)	194 (60.8)
Nonbinary	7 (4.4)	11 (6.8)	18 (5.6)
I don’t wish to say	2 (1.3)	2 (1.2)	4 (1.3)
State, n (%)			
Australian Capital Territory	1 (0.6)	2 (1.2)	3 (0.9)
New South Wales	63 (40.1)	65 (40.1)	128 (40.1)
Northern Territory	0 (0)	1 (0.6)	1 (0.3)
Queensland	17 (10.8)	13 (8)	30 (9.4)
South Australia	10 (6.4)	10 (6.2)	20 (6.3)
Tasmania	1 (0.6)	3 (1.9)	4 (1.2)
Victoria	35 (22.3)	36 (22.2)	71 (22.3)
Western Australia	29 (18.5)	27 (16.7)	56 (17.6)
I don’t wish to say	1 (0.7)	5 (3.1)	6 (1.9)
Remoteness, n (%)			
Nonmajor city	12 (7.7)	13 (8)	25 (7.8)
Major city	117 (74.5)	120 (74.1)	237 (74.3)
I don’t wish to say	28 (17.8)	29 (17.9)	57 (17.9)
Language other than English spoken at home, n (%)			
No	99 (63.1)	95 (58.6)	194 (60.8)
Yes	58 (36.9)	61 (37.7)	119 (37.3)
I don’t wish to say	0 (0)	6 (3.7)	6 (1.9)
Aboriginal or Torres Strait Islander, n (%)			
No	152 (96.8)	153 (94.4)	305 (95.6)
Yes	2 (1.3)	4 (2.5)	6 (1.9)
I don’t wish to say	3 (1.9)	5 (3.1)	8 (2.5)
Currently studying, n (%)			
No	3 (1.9)	9 (5.5)	12 (3.8)
Yes	151 (96.2)	148 (91.4)	299 (93.7)
I don’t wish to say	3 (1.9)	5 (3.1)	8 (2.5)
Current or highest level of study, n (%)			
High school	104 (66.2)	107 (66.1)	211 (66.2)
Post high school	50 (31.9)	49 (30.2)	99 (31)
I don’t wish to say	3 (1.9)	6 (3.7)	9 (2.8)
School type, n (%)			
Public	108 (68.8)	102 (63)	210 (65.8)
Catholic	17 (10.8)	28 (17.3)	45 (14.1)
Private	31 (19.8)	26 (16.1)	57 (17.9)
I don’t wish to say	1 (0.6)	6 (3.7)	7 (2.2)
Sexual identity, n (%)			

XSL•FO
RenderX

Variables	Scaled (n=157)	Fixed (n=162)	Total (n=319)
Heterosexual	103 (65.6)	106 (65.5)	209 (65.5)
LGBTQIA[a]	53 (33.8)	54 (33.3)	107 (33.5)
I don’t wish to say	1 (0.6)	2 (1.2)	3 (1.0)
Recruitment setting, n (%)			
Instagram/Facebook	17 (10.8)	20 (12.4)	37 (11.6)
Student Edge	140 (89.2)	142 (87.6)	282 (88.4)
Recruitment year, n (%)			
2022	63 (40.1)	58 (35.8)	121 (37.9)
2023	94 (59.9)	104 (64.2)	198 (62.1)

[a]LGBTQIA: lesbian, gay, bisexual, transgender, queer, intersex, and asexual.

Primary Analyses

Overview

Participants nominated an average of 3.47 (IQR 2 - 5, range 0 - 10) peers. Peers who were ineligible based on age were excluded (n=4). This left 298 (of 319) participants who had nominated at least one eligible peer. Of 1069 total peers nominated, 55 completed the survey, of which 48 (aged 16-21 y) were successfully linked with their seed. The remaining peers were unable to be linked via the ID number of their referring seed and peer initials or codename. In total, 11.9% (n=38) of seeds successfully recruited a peer (18.5% scaled and 5.6% fixed). Only participants in the scaled group recruited more than one peer (3.8% recruited two peers and 1.3% recruited three peers).

Primary Outcome

Those in the scaled reimbursement group were approximately 3.80 times more likely to successfully recruit their nominated peers than those in the fixed reimbursement group (proportion ratio 3.80, 95% CI 1.78 - 8.09).

Secondary Outcome

Results from the Poisson regression analysis showed an effect of intervention group on the number of peers recruited. Each participant in the scaled group recruited an average of 0.25 peers, while those in the fixed group recruited an average of 0.06 peers, resulting in a difference of 0.19 (95% CI 0.11 - 0.28) between groups.

Secondary Analyses

Primary Outcome

Results from exploratory bivariable log binomial GEE analyses can be found in Table 2. The proportion of nominated peers successfully recruited was similar regardless of the gender, age, education level, and network size of seeds, or the gender, age, and closeness of peers. On average, seeds recruited through Instagram or Facebook were approximately two times as likely to successfully recruit their nominated peers. However, the CI was wide, indicating the difference between seeds recruited through Instagram or Facebook and those recruited through Student Edge was relatively imprecise (proportion ratio 2.20, 95% CI 1.06 - 4.55).

XSL•FO
RenderX

Table . Seed and peer-level factors associated with successful recruitment of a nominated peer.

Variables	PR[a] (95% CI)
Seed level	
Gender	
Man	—[b]
Woman	1.04 (0.52-2.09)
Nonbinary	1.82 (0.61-5.31)
Age (years)	
16	—
17	0.83 (0.37-1.85)
18	1.02 (0.48-2.18)
Education	
High school	—
Post high school	1.05 (0.54-2.06)
Recruitment method	
Student Edge	—
Instagram/Facebook	2.20 (1.06-4.55)
Network size	
≤5 peers	—
>5 peers	0.46 (0.17-1.21)
Peer level	
Gender	
Man	—
Woman	0.89 (0.48-1.67)
Nonbinary	1.93 (0.75-4.97)
Age (years)	
16	—
17	0.56 (0.28-1.13)
18	0.50 (0.22-1.15)
≥19	0.98 (0.41-2.38)
Closeness	
Not at all/slightly/moderately	—
Very/extremely	1.13 (0.62-2.05)

[a]PR: proportion ratio.

[b]Not applicable.

Secondary Outcome

Results from bivariable Poisson regression analyses using seed-level variables can be found in Table 3. The number of nominated peers successfully recruited was similar regardless of the gender, age, education level, and network size of seeds. On average, seeds recruited through Instagram or Facebook were likely to recruit approximately two and a half times more peers. However, the wide CI indicates that the estimate of this difference between seeds recruited through Instagram or Facebook and those recruited through Student Edge was relatively imprecise (count ratio 2.54, 95% CI 1.32 - 4.88).

XSL•FO
RenderX

Table . Seed-level factors associated with recruiting a larger number of nominated peers.

Variables	CR[a] (95% CI)
Gender	
Man	—[b]
Woman	0.99 (0.53-1.86)
Nonbinary	1.91 (0.69-5.25)
Age (years)	
16	—
17	0.93 (0.45-1.89)
18	1.08 (0.54-2.15)
Educated	
High school	—
Post high school	0.97 (0.53-1.78)
Recruitment method	
Student Edge	—
Instagram/Facebook	2.54 (1.32-4.88)
Network size	
≤5 peers	—
>5 peers	1.13 (0.53-2.41)

[a]CR: count ratio.

[b]Not applicable.

Reimbursement Costs

Excluding additional costs associated with using the student panel or advertising the study on social media, overall, the average reimbursement per participant was AUD $6.39 (US $4.20). Separated by group, average reimbursement per participant was AUD $5 (US $3.30) in the fixed intervention and AUD $7.60 (US $5.00) in the scaled condition. Specifically, an average seed received AUD $5.61 (US $3.68; fixed group AUD $5, US $3.29; scaled group AUD $6.24, US $4.10), and an average peer received AUD $11.56 (US $7.60; fixed group AUD $5, US $3.29; scaled group AUD $13.08, US $8.60).

Discussion

Principal Results

As predicted, a scaled reimbursement strategy that included additional incentives for each peer in a network was more effective than a fixed reimbursement strategy in recruiting adolescent peer networks. Those in the scaled reimbursement group recruited a greater proportion of nominated peers and recruited a greater average number of peers, compared to the fixed reimbursement group. Additionally, seeds recruited through Instagram and Facebook were more likely to recruit peers and each recruit a higher number of peers than those recruited through Student Edge. Other demographic and relational factors of seeds or peers were not found to have an impact on peer recruitment success.

Comparison With Prior Work

While we are not aware of any studies to date that have compared the relative effectiveness of different financial incentives for recruiting adolescent peers, our results are consistent with those from a WebRDS study with a sample of men who have sex with men, which identified a preference for scaled over fixed reimbursement for recruiting peers in their network [19]. More broadly, they also reflect the findings of a review that reported lower recruitment success among WebRDS studies with no monetary incentive [11]. The study by Schonlau et al [18] is also relevant, as it experimentally varied reimbursement amounts for each peer recruited (US $15 vs $30) and the inclusion of sign-up bonuses for peers (US $0 vs $20) to determine the most effective way to recruit adults into an internet panel. The combination of higher incentives and a sign-up bonus significantly increased peer recruitment compared to either one alone [18]. Despite key differences between that study and the current study, this highlights the importance of adequate monetary reimbursements for both seeds and peers. In our study, the additional cost for scaled reimbursements was AUD $2.60 (US $1.68) per participant, which is minimal when considered against the total expenses of conducting the study. Although peer recruitment in the scaled condition was improved compared to the fixed condition (18.5% vs 5.6%), overall peer recruitment in this study was low (11.9%). We conservatively estimated recruiting 92 peers, yet only obtained 55, not all of whom could be linked automatically to their seed due to ethical constraints. Despite allowing participants to recruit up to 10 peers each and incentives scaling up to the fifth peer for those in the scaled group, very few seeds recruited more than one

XSL•FO
RenderX

peer. Similar studies conducted with youth samples have reported peer recruitment rates by seeds of 61.8% and 66.7%, although these studies had fewer initial seeds, higher incentives, and capped referrals at three per participant. Helms et al [11] highlighted that more initial seeds tended to result in less successfully recruited peers, possibly due to fewer resources being allocated to engage seeds or seek out seeds with a larger social network. This may explain the low rates of recruitment in our study, where seeds were recruited online (via a research panel for students or social media advertisements) and informed of the study requirements at the beginning of the survey rather than directly through research staff. A drawback of offering monetary incentives to participants is the increased potential of recruiting those who are not inherently engaged with the topic and therefore not motivated to refer their peers [10]. Since most participants were recruited from an online panel of students (Student Edge) who are used to completing short surveys in exchange for small payments, reaching out to peers to complete the survey—without the guarantee that they would finish it and thus receive extra payment—might not have been very appealing. This may also explain the finding from exploratory analyses that seeds recruited through social media were more likely to recruit their peers than those recruited through Student Edge.

Although recruitment through Student Edge allowed us to quickly recruit the remainder of our sample, it did not result in effective peer recruitment. Few WebRDS studies in the literature have used research panels to recruit peers. Stein et al [28] sampled seeds from two internet participatory surveillance panels comprising a community of adult volunteers who regularly report on health symptoms. While those in the panel with the chance to win a gift card were more likely to recruit peers, overall, both samples successfully recruited peers [28]. Notably, participants from that study likely had different motivations for engaging in the research (ie, interest in assisting in health surveillance) than our participants recruited through the Student Edge panel.

Although social media recruitment was more effective for recruiting peers, no seeds were recruited through Snapchat or TikTok, and only a small number through Instagram and Facebook. Currently, WebRDS studies have largely recruited young people and adults using Facebook, and there are few examples of researchers using any platforms other than Meta (Instagram and Facebook) [12,14,29]. Considering the key differences between social media platforms, video content should be tailored to the platform. For instance, our advertisements comprised short videos with colorful and dynamic text, yet TikTok recruitment videos typically feature an individual (ie, health professional or peer) introducing the research [30]. There is no clear formula for success in social media recruitment. There are many factors that contribute to recruitment success, such as ad style and content, advertising budget, the account that distributes the ad, and interactions between users and the ads [31]. While many studies have reported on successful recruitment and made recommendations such as use of colorful images, pop-culture memes, or bold text, other studies attempting to duplicate these methods fall short [31-34].

Consistent with other WebRDS research, successful and nonsuccessful seeds did not differ in gender, age [35,36], or education level [36]. Although some general population studies have reported that successful seeds are more likely to be younger (19-24 y), our age range was quite narrow (16 - 18 y), suggesting that there may be no meaningful difference within these later teen years. Similarly, recruited peers did not differ in age, gender, or closeness of relationship with their referring seed, relative to nonrecruited peers. To our knowledge, no studies have directly examined this, although many have reported on homophily (the tendency for seeds to recruit peers with similar characteristics) present within samples [10]. Theoretically, it would be useful to understand relationship factors driving peers to engage in research (ie, if closeness of relationship was predictive of engagement, we could encourage seeds to only target these relationships). Nonetheless, most referred peers in our study had very (40.2%) or extremely (28.5%) close connections, suggesting that seeds are more inclined to recruit closer friends anyway. Overall, these findings emphasize the importance of focusing on financial reimbursement and recruitment source when attempting to recruit adolescent peer networks.

Strengths and Limitations

Our study compared the effectiveness of two reimbursement strategies in a WebRDS study with adolescents aged 16 to 18 years in an RCT. We experimented using social media advertisements and a student panel to recruit young people, expanding beyond school-based relationships. The use of two-factor authentication allowed us to largely address the issue of fraudulent responses commonly reported in the WebRDS literature [10]. Nonetheless, we encountered other methodological issues, namely around the effectiveness of recruitment sources and recruitment of peers. Issues recruiting young people through social media resulted in a delay in recruitment while advertisements were revised. As a result, recruitment took place in two phases over a 2-year period. This does not appear to have significantly impacted the nature of the sample, as there were no key demographic differences between the two recruitment phases. A further deviation from protocol occurred because we could not reach a sufficient sample size from each recruitment source, and most participants were recruited through a market research panel. We have previously found that participants recruited through survey panels demonstrated greater attrition in a longitudinal study compared to those recruited through social media [37]. Another study showed that a research panel was significantly more representative than a sample recruited via social media [38]. It is recommended that future studies invest sufficient resources into creating advertising content that is engaging and relevant for specific social media platforms.

We were able to recruit sufficient seeds into each group, but the peer sample was smaller than anticipated based on previous literature. While the sample was sufficient to answer the primary research question, showing that scaled incentives are more effective than fixed incentives for participant recruitment, the smaller than anticipated number of peers recruited in both study arms raises questions about whether further methodological adjustments are needed to maximize peer recruitment. We found

XSL•FO
RenderX

that those recruited through social media were more likely to refer peers than those recruited through the Student Edge panel. Further research is warranted to determine why that may be the case. As a result, the confidence intervals for exploratory analyses were wide. As such, results from secondary analyses should be interpreted with caution.

Notably, seven peers were unable to be linked with their seed using their codename, initials, and their seed's ID number, highlighting a limitation of relying on manual linking methods. Due to the higher attrition rate post unblinding in the fixed intervention group, it is possible that participants stopped the survey when advised of their intervention group, which may overestimate the effectiveness of the intervention.

Finally, this study was conducted with a specific population of adolescents obtained through convenience sampling; hence, the sample is not representative of the population, and findings may not translate to other age groups and samples. Compared to Australian census data, our sample included a higher proportion identifying as lesbian, gay, bisexual, transgender, queer, intersex, asexual, and other identities (34% vs 10%), speaking a language other than English at home (37% vs 23%), identifying as Aboriginal and Torres Strait Islander (3.2% vs 1.9%), and attending government schools (65% vs 59%) [39].

Conclusions

The present findings highlight important considerations for using WebRDS methodology to recruit adolescent peer networks. Specifically, adolescents are more motivated to recruit their peers to complete a survey when incentivized for each additional peer, although this does not guarantee a large peer network. Participants recruited through Instagram and Facebook were more likely to recruit their peers than those from the student panel, possibly due to different motivations for research involvement. However, peer networks remained small across both recruitment methods. Further research is warranted to understand how to best recruit adolescent peer networks outside of the school setting using financial reimbursements and different recruitment sources. With more time and cost-effective web-based approaches to recruiting peers, longitudinal studies can be conducted to investigate the influence of social networks on key health behaviors over time.

Funding

This study received no specific funding. MR was supported by a Monash University Training Program scholarship. RSD, PD, MH, and JH were supported by National Health and Medical Research Council Investigator grants. MSCL was supported by National Health and Medical Research Council Career Development Fellowship.

Disclaimer

Artificial intelligence was not used in any part of this manuscript writing.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to participant privacy but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: MR, CW, PD, MH, MSCL
Methodology: RSD, MR, CW, JH, MSCL
Data collection: SE, MR, CW, MSCL
Analysis: SE, RSD, MR
Writing: SE, MR, MSCL
Review and Editing: all authors

Conflicts of Interest

None declared.

Checklist 1
CONSORT-eHEALTH (V 1.6.1) checklist.
[PDF File, 951 KB - i-jmr_v14i1e70223_app1.pdf]

References

1. Viner RM, Ozer EM, Denny S, et al. Adolescence and the social determinants of health. The Lancet 2012 Apr;379(9826):1641-1652. [doi: 10.1016/S0140-6736(12)60149-4]
2. Garrido M, Sufrinko N, Max J, Cortes N. Where youth live, learn, and play matters: tackling the social determinants of health in adolescent sexual and reproductive health. Am J Sex Educ 2018 Jul 3;13(3):269-282. [doi: 10.1080/15546128.2018.1458264]

XSL•FO
RenderX

3. Valente TW. Social networks and health behavior. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior: Theory, Research, and Practice, 5th edition: Jossey-Bass/Wiley: Hoboken; 2015:205-222.
4. Montgomery SC, Donnelly M, Bhatnagar P, Carlin A, Kee F, Hunter RF. Peer social network processes and adolescent health behaviors: a systematic review. Prev Med 2020 Jan;130:105900. [doi: 10.1016/j.ypmed.2019.105900] [Medline: 31733224]
5. Leung RK, Toumbourou JW, Hemphill SA. The effect of peer influence and selection processes on adolescent alcohol use: a systematic review of longitudinal studies. Health Psychol Rev 2014;8(4):426-457. [doi: 10.1080/17437199.2011.587961] [Medline: 25211209]
6. Baams L, Overbeek G, van de Bongardt D, Reitz E, Dubas JS, van Aken MAG. Adolescents' and their friends' sexual behavior and intention: selection effects of personality dimensions. J Res Pers 2015 Feb;54:2-12. [doi: 10.1016/j.jrp.2014.07.009]
7. Jeon KC, Goodson P. US adolescents' friendship networks and health risk behaviors: a systematic review of studies using social network analysis and Add Health data. PeerJ 2015;3:e1052. [doi: 10.7717/peerj.1052] [Medline: 26157622]
8. Gmel G, Marmet S, Wicki M, Studer J, Dietze P. The association between changes in peer pressure and changes in alcohol and cannabis use: a prospective cohort study among men during emerging adulthood. Eur Addict Res 2025 Jul 16;31(2):113-124. [doi: 10.1159/000544858]
9. Valente TW, Pitts SR. An appraisal of social network theory and analysis as applied to public health: challenges and ppportunities. Annu Rev Public Health 2017 Mar 20;38(1):103-118. [doi: 10.1146/annurev-publhealth-031816-044528] [Medline: 27992729]
10. Ferrer-Rosende P, Feijoo-Cid M, Fernández-Cano MI, Salas-Nicás S, Stuardo-Ávila V, Navarro-Giné A. Implementation of web-based respondent driven sampling in epidemiological studies. BMC Med Res Methodol 2023 Oct 2;23(1):217. [doi: 10.1186/s12874-023-02042-z] [Medline: 37784020]
11. Helms YB, Hamdiui N, Kretzschmar MEE, et al. Applications and recruitment performance of web-based respondent-driven sampling: scoping review. J Med Internet Res 2021 Jan 15;23(1):e17564. [doi: 10.2196/17564] [Medline: 33448935]
12. Hildebrand J, Burns S, Zhao Y, et al. Potential and challenges in collecting social and behavioral data on adolescent alcohol norms: comparing respondent-driven sampling and web-based respondent-driven sampling. J Med Internet Res 2015 Dec 24;17(12):e285. [doi: 10.2196/jmir.4762] [Medline: 26704736]
13. Liu H, Li J, Han M, et al. Optimizing respondent-driven sampling survey to identify factors associated with HIV risk behaviors among student MSM aged 15 to 19. Inquiry 2023;60:469580231166811. [doi: 10.1177/00469580231166811] [Medline: 37042211]
14. Bauermeister JA, Zimmerman MA, Johns MM, Glowacki P, Stoddard S, Volz E. Innovative recruitment using online networks: lessons learned from an online study of alcohol and other drug use utilizing a web-based, respondent-driven sampling (webRDS) strategy. J Stud Alcohol Drugs 2012 Sep;73(5):834-838. [doi: 10.15288/jsad.2012.73.834] [Medline: 22846248]
15. Wejnert C, Heckathorn DD. Web-based network sampling: efficiency and efficacy of respondent-driven sampling for online research. Sociol Methods Res 2008;37(1):105-134. [doi: 10.1177/0049124108318333]
16. Tucker JA, Bacon JP, Chandler SD, Lindstrom K, Cheong J. Utility of digital respondent driven sampling to recruit community-dwelling emerging adults for assessment of drinking and related risks. Addict Behav 2020 Nov;110:106536. [doi: 10.1016/j.addbeh.2020.106536] [Medline: 32711287]
17. Abdelazeem B, Abbas KS, Amin MA, et al. The effectiveness of incentives for research participation: a systematic review and meta-analysis of randomized controlled trials. PLOS ONE 2022;17(4):e0267534. [doi: 10.1371/journal.pone.0267534] [Medline: 35452488]
18. Schonlau M, Weidmer B, Kapteyn A. Recruiting an internet panel using respondent-driven sampling. J Off Stat 2014 Jun 1;30(2):291-310. [doi: 10.2478/jos-2014-0018]
19. Diexer S, Teslya A, Buskens V, Matser A, Stein M, Kretzschmar ME. Improving web-based respondent-driven sampling performance among men who have sex with men in The Netherlands. PLOS Digit Health 2023 Feb;2(2):e0000192. [doi: 10.1371/journal.pdig.0000192] [Medline: 36812647]
20. Sang JM, Gholamian B, Wang L, et al. Examining differential success in recruitment using respondent driven sampling (RDS) in a multi-site study of gay, bisexual and other men who have sex with men. BMC Med Res Methodol 2023 Jun 9;23(1):136. [doi: 10.1186/s12874-023-01886-9] [Medline: 37296373]
21. Kuhns LM, Kwon S, Ryan DT, Garofalo R, Phillips G 2nd, Mustanski BS. Evaluation of respondent-driven sampling in a study of urban young men who have sex with men. J Urban Health 2015 Feb;92(1):151-167. [doi: 10.1007/s11524-014-9897-0] [Medline: 25128301]
22. Raggatt M, Wright CJC, Sacks-Davis R, et al. Identifying the most effective recruitment strategy using financial reimbursements for a web-based peer network study with young people aged 16-18 years: protocol for a randomized controlled trial. JMIR Res Protoc 2023 Aug 11;12:e44813. [doi: 10.2196/44813] [Medline: 37566448]
23. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126. [doi: 10.2196/jmir.1923] [Medline: 22209829]

XSL•FO
RenderX

24. Student Edge. URL: https://studentedge.org/au [accessed 2024-12-03]
25. 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]
26. de la Haye K, Robins G, Mohr P, Wilson C. Obesity-related behaviors in adolescent friendship networks. Soc Networks 2010 Jul;32(3):161-167. [doi: 10.1016/j.socnet.2009.09.001]
27. Mercken L, Snijders TAB, Steglich C, de Vries H. Dynamics of adolescent friendship networks and smoking behavior: social network analyses in six European countries. Soc Sci Med 2009 Nov;69(10):1506-1514. [doi: 10.1016/j.socscimed.2009.08.003]
28. Stein ML, van Steenbergen JE, Buskens V, et al. Enhancing syndromic surveillance with online respondent-driven detection. Am J Public Health 2015 Aug;105(8):e90-e97. [doi: 10.2105/AJPH.2015.302717] [Medline: 26066940]
29. Sadasivam RS, Cutrona SL, Luger TM, et al. Share2Quit: online social network peer marketing of tobacco cessation systems. Nicotine Tob Res 2017 Mar 1;19(3):314-323. [doi: 10.1093/ntr/ntw187] [Medline: 27613918]
30. Lee N, Hong Y, Hu S, Kirkpatrick CE, Lee S, Hinnant A. Exploring the strategic use of TikTok for clinical trial recruitment: how audiences' prior short-form video usage influences persuasive effects. J Health Commun 2024 Apr 2;29(4):294-306. [doi: 10.1080/10810730.2024.2339237] [Medline: 38590176]
31. Coombe J, Bittleston H, Ludwick T, et al. Recruiting participants via social media for sexual and reproductive health research. Sex Health 2025 Jun;22:22. [doi: 10.1071/SH24123] [Medline: 40526809]
32. Ford KL, Albritton T, Dunn TA, Crawford K, Neuwirth J, Bull S. Youth study recruitment using paid advertising on Instagram, Snapchat, and Facebook: cross-sectional survey study. JMIR Public Health Surveill 2019 Oct 9;5(4):e14080. [doi: 10.2196/14080] [Medline: 31599739]
33. Darko EM, Kleib M, Olson J. Social media use for research participant recruitment: integrative literature review. J Med Internet Res 2022 Aug 4;24(8):e38015. [doi: 10.2196/38015] [Medline: 35925655]
34. Lim MSC, Molenaar A, Brennan L, Reid M, McCaffrey T. Young adults' use of different social media platforms for health information: insights from web-based conversations. J Med Internet Res 2022 Jan 18;24(1):e23656. [doi: 10.2196/23656] [Medline: 35040796]
35. Jonsson J, Stein M, Johansson G, Bodin T, Strömdahl S. A performance assessment of web-based respondent driven sampling among workers with precarious employment in Sweden. PLoS ONE 2019;14(1):e0210183. [doi: 10.1371/journal.pone.0210183] [Medline: 30629661]
36. Theunissen K, Hoebe C, Kok G, et al. A web-based respondent driven sampling pilot targeting young people at risk for Chlamydia Trachomatis in social and sexual networks with testing: a use evaluation. Int J Environ Res Public Health 2015 Aug 20;12(8):9889-9906. [doi: 10.3390/ijerph120809889] [Medline: 26308015]
37. Orozco A, Thomas A, Raggatt M, et al. Coping with COVID-19: a prospective cohort study on young Australians' anxiety and depression symptoms from 2020-2021. Arch Public Health 2024 Sep 26;82(1):166. [doi: 10.1186/s13690-024-01397-z] [Medline: 39327590]
38. Lehdonvirta V, Oksanen A, Räsänen P, Blank G. Social media, web, and panel surveys: using non‐probability samples in social and policy research. Policy Internet 2021 Mar;13(1):134-155. [doi: 10.1002/poi3.238]
39. 4221.0 - schools, australia. Australian Bureau Statistics. 2016. URL: https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/4221.0Main+Features12016 [accessed 2025-11-12]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
GEE: generalized estimating equations
RCT: randomized controlled trial
RDS: respondent-driven sampling
WebRDS: web-based respondent-driven sampling

Edited by N Cahill, TDA Cardoso; submitted 17.12.24; peer-reviewed by D Marshall, M Hecht; revised version received 19.10.25; accepted 22.10.25; published 25.11.25.

Please cite as:
Eddy S, Sacks-Davis R, Raggatt M, Wright C, Dietze P, Hellard M, Hocking J, Lim MSC
Enhancing Recruitment of Adolescents Aged 16-18 Years in a Web-Based Peer Network Study Through Financial Reimbursements: Randomized Controlled Trial
Interact J Med Res 2025;14:e70223
URL: https://www.i-jmr.org/2025/1/e70223
doi:10.2196/70223

XSL•FO
RenderX

© Sarah Eddy, Rachel Sacks-Davis, Michelle Raggatt, Cassandra Wright, Paul Dietze, Margaret Hellard, Jane Hocking, Megan S C Lim. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 25.11.2025. 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

Increasing Participation and Completion Rates in Questionnaire Surveys of Primary Care Patients: Cluster-Randomized Study

Paul Sebo[1], MD, MSc; 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]University Institute for Primary Care (IuMFE), University of Geneva, Geneva, Switzerland

[2]University College of General Medicine, Université Claude Bernard Lyon 1, Lyon, France

Corresponding Author:
Paul Sebo, MD, MSc
University Institute for Primary Care (IuMFE)
University of Geneva
Rue Michel-Servet 1
Geneva, 1211
Switzerland
Phone: 41 22 379 50 61
Email: paulsebo@hotmail.com

Abstract

Background: Participation and completion rates in questionnaire-based surveys are often low.

Objective: This study aims to assess participation and completion rates for a survey using paper and mixed mode questionnaires with patients recruited by research assistants in primary care waiting rooms.

Methods: This cluster-randomized study, conducted in 2023 in France, involved 974 patients from 39 practices randomized into 4 groups: “paper with incentive” (n=251), “paper without incentive” (n=368), “mixed mode with tablet” (n=187), and “mixed mode with QR code” (n=168). Analyses compared the combined paper group with the 2 mixed mode groups and the “paper with incentive” and “paper without incentive” groups. Logistic regressions were used to analyze participation and completion rates.

Results: Of the 974 patients recruited, 822 (women: 536/821, 65.3%; median age 52, IQR 37-68 years) agreed to participate (participation rate=84.4%), with no significant differences between groups. Overall, 806 patients (98.1%) answered all 48 questions. Completion rates were highest in the combined paper group (99.8%) compared to mixed mode groups (96.8% for paper or tablet, 93.3% for paper or QR code; P<.001). There was no significant difference in completion rates between the “paper with incentive” and “paper without incentive” groups (100% vs 99.7%).

Conclusions: Recruiting patients in waiting rooms with research assistants resulted in high participation and completion rates across all groups. Mixed mode options did not enhance participation or completion rates but may offer logistical advantages. Future research should explore incentives and mixed-mode strategies in diverse settings.

(Interact J Med Res 2025;14:e67981) doi:10.2196/67981

KEYWORDS

completion rate; missing data; mixed mode; web-based; participation rate; primary care; questionnaire; QR code; tablet; survey; primary care patients; randomized study

Introduction

Questionnaire-based surveys are valuable tools in a variety of research areas, including health care. For example, they allow data to be collected on patients’ experiences, opinions, and behaviors, which helps health care professionals to better understand their needs and improve the quality of care. Apart from the fact that they are generally inexpensive and take little time, 1 major advantage of questionnaire-based surveys lies in the flexibility they can offer in terms of the mode of administration (ie, face-to-face interviews, telephone interviews, and paper or web-based questionnaires completed at home or elsewhere).

However, to maintain a good level of representativeness and avoid selection bias, it is crucial to obtain high participation and completion rates. For surveys with postal, email, or telephone recruitment, several studies showed that participation rates (especially for web-based questionnaires) [1-4] and

XSL•FO
RenderX

completion rates (especially for paper questionnaires) [1,2] tended to be relatively low, whether the participants were individuals, patients, or physicians. Other studies showed that physicians' participation and completion rates were generally lower than those of the general population, probably mainly because of their workload in the practice and the increasing frequency with which they were asked to respond to surveys [5,6]. Although participation and completion rates are generally particularly low among physicians, these rates may also be suboptimal in studies with patients [7]. This means that the results of a large number of surveys conducted with patients may not be representative of the target population as a whole and may therefore lead to misleading conclusions.

Among the factors likely to influence participation and completion rates, the method chosen to collect data and the use of incentives are key elements [8-12]. Mixed modes, that is, the use of several methods in the same study, seem particularly useful for improving participation rates [9,13]. To our knowledge, QR codes have only been the subject of a few studies, focusing on specific populations and with limited sample sizes [11,14-16]. The usefulness of mixed modes including QR codes to improve participation and completion rates in a questionnaire has not yet been explored in general practice.

The aim of this study was to compare the participation and completion rates for a 48-question survey across 4 groups: a paper questionnaire group with incentive, a paper questionnaire group without incentive, and 2 mixed mode groups (ie, paper or web-based with tablet, and paper or web-based via QR code). Specifically, we analyzed differences between the combined paper group and the 2 mixed mode groups, as well as between the "paper with incentive" and "paper without incentive" groups. Patients were recruited into primary care waiting rooms by research assistants who were available to answer their questions if necessary. We hypothesized that both mixed mode options and incentives would improve participation and completion rates.

Methods

Design, Setting, and Study Population

This cluster-randomized study carried out in 2023 in the Rhône-Alpes region (France) was part of an environmental health project aimed at profiling different patterns of meat consumers in primary care, with a view to designing brief interventions to reduce meat consumption.

We used a professional register of primary care physicians from which we randomly extracted 200 physicians using computer-generated random numbers. Five research assistants contacted each randomly selected primary care physician by email until the required number of participating physicians (ie, n=40, Sample Size Determination section) was reached. If the primary care physician refused to participate or did not respond after 3 consecutive reminders, we contacted the next practice on the list.

The research assistants in this study were postgraduate residents in primary care medicine, conducting this work as part of their MD thesis. They had prior clinical experience, which enabled them to effectively interact with patients. To ensure consistency in patient recruitment and minimize variability, all research assistants underwent standardized training before the study began. This training included guidance on presenting the study, explaining the information sheet, and following ethical protocols. Although the study covered a large geographic area, requiring assistants to work autonomously during the recruitment phase, their approach was standardized and aligned with the study's protocols. Due to logistical constraints, the assistants were not randomly assigned to specific clinics.

We used simple randomization to allocate our sample of medical practices into 4 groups using computer-generated random numbers. Fifteen practices were randomized into the "paper without incentive" group, 10 into the "paper with incentive" group, 7 into the "paper or web-based with tablet" group (ie, patients were given the option of using either paper or a tablet provided by the researchers), and 7 into the "paper or web-based via QR code" group (ie, patients were given the option of using either paper or a QR code if they had their smartphones with them). We used computer-generated random numbers to determine the day of the week on which the study would be carried out in each practice.

Ethical Considerations

The study was approved by the Research Ethics Committee of the University College of General Practice, Claude Bernard Lyon 1 University (2023-01-03-01). The committee granted approval for the study protocol, including patient recruitment and data collection procedures. No additional ethical approvals were required beyond this institutional review. Participants were nonurgent, French-speaking, consecutive adult patients who were able to understand the study and provide written informed consent. They were informed about the study's objectives, data confidentiality, and their right to withdraw at any time without consequences. No personally identifiable information was collected, and all data were anonymized before analysis to ensure participant privacy and confidentiality. In line with the environmental health theme of the study, an origami paper containing a seed was provided as an unconditional incentive to all patients in the "paper with incentive" group, regardless of their willingness to complete the questionnaire. No financial compensation was offered to participants. No identifiable features of research participants are present in any images or supplementary materials included in the manuscript.

Data Collection

Participants were informed by a poster and recruited in the waiting room by a research assistant (20-25 patients per practice) between January 9, 2023, and June 16, 2023. The research assistant verified that the inclusion criteria were met and was available to answer any questions from the participants. The questionnaire (Multimedia Appendix 1) consisted of 48 questions: 4 sociodemographic questions (age, sex, postcode, and occupation), the French version of the feeling of the Sense of Coherence Scale (=17 questions), the French version of the Meat Attachment Questionnaire (=17 questions), and a questionnaire on intentionality adapted from previous studies (=10 questions) [17-19]. The questionnaire generally took about 10 minutes to complete. The paper and web-based questionnaires

XSL•FO
RenderX

were designed to be as similar as possible. In the 2 mixed modes (ie, with a tablet or QR code), we included an alert indicating that one or more questions were not answered. However, to be consistent with the paper version, participants did not have to answer all the questions before submitting the web-based questionnaire.

Statistical Analyses and Sample Size Determination

The sample size calculation was conducted to ensure sufficient power to detect differences in completion rates between groups. For simplification, the calculation was based on comparing proportions between 2 groups. Completion rates reported in the literature vary widely, ranging from 35% to 99%, depending on factors such as the study population, the theme of the study, and the format of the questionnaire [1,2,4]. We hypothesized that the completion rate in our study would be at least 80%, given the presence of a research assistant in the waiting room to assist participants if needed.

We calculated that 428 patients would be required in total to detect a 10-percentage-point difference between groups, with a power of 80% and a significance level of 5%, prior to accounting for clustering [20]. To adjust for clustering within practices, we assumed a conservative intraclass correlation coefficient of 0.02, based on findings from prior research in primary care settings [21], and an average cluster size of 20 patients per practice. This resulted in a design effect of 1.38, inflating the total adjusted sample size to 591 patients. With an average cluster size of 20 patients, this corresponded to 30 clusters in total. As 3 groups were compared, the sample size was increased to 40 clusters to ensure sufficient power.

We computed the proportion of patients who agreed to take part in the study (participation rate), the median number of questions answered with IQR, and the proportion of patients who answered all the questions (completion rate).

To assess whether the participation rate differed between groups, we used logistic regressions adjusted for intracluster correlation within practices. Comparisons were made between the combined paper group, the “paper or tablet” group, and the “paper or QR code” group as well as between the “paper with incentive” and “paper without incentive” groups. Multivariable analyses were not conducted for participation rates, as sociodemographic data were not collected for patients who refused to participate, in accordance with the ethics committee’s recommendations. This limitation prevented us from assessing variations in participation rates by sociodemographic factors.

For the completion rate, exact logistic regressions adjusted for intracluster correlation were used to address sparse data issues, where cells formed by the outcome and categorical predictor variables had few or no observations [22]. Comparisons were made for (1) the three groups based on initial randomization (ie, combined paper, “paper or tablet,” and “paper or QR code”); (2) groups based on the questionnaire format chosen (ie, paper, “web-based with tablet,” and “web-based with QR code”); (3) the “paper with incentive” and “paper without incentive” groups; and (4) sex and age group (<40, 40-60, and ≥60). Multivariable analyses were also carried out, adjusting for sex, age group, and occupation.

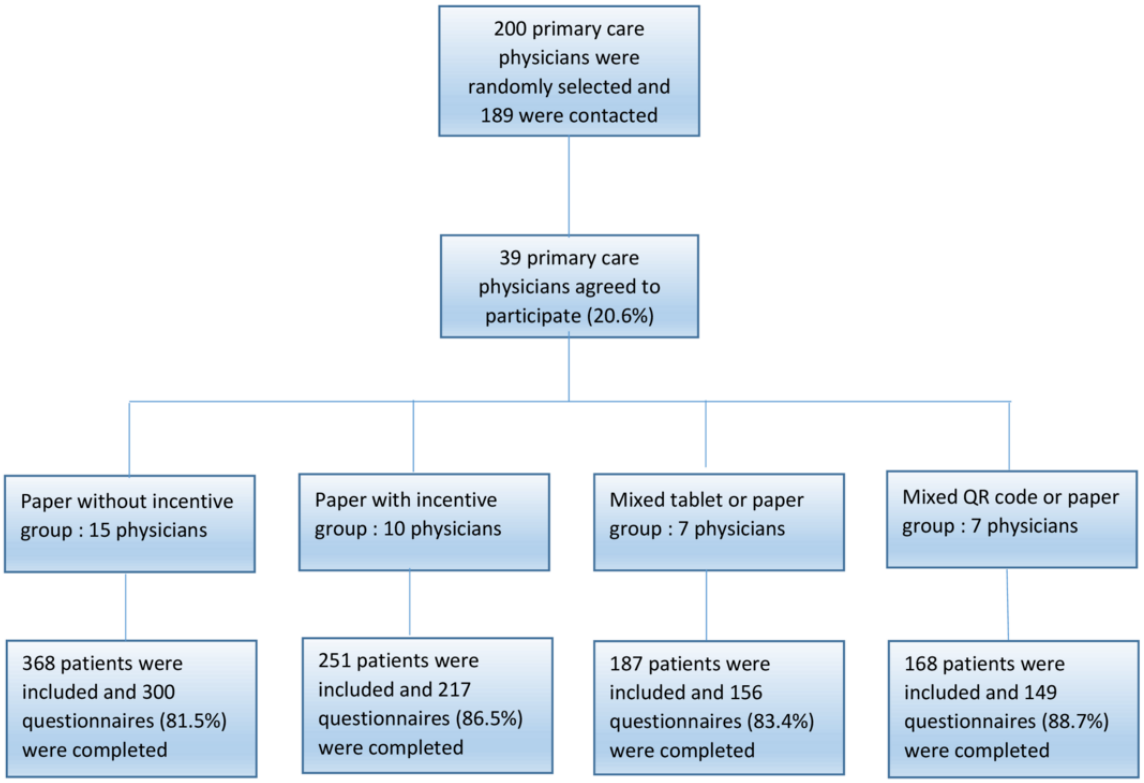
All analyses were carried out with STATA (version 15.1; StataCorp LLC).

Results

The flowchart for the study is shown in Figure 1. From the list of 200 primary care physicians, 189 were contacted, of whom 39 (20.6%) agreed to participate. The study included 15 primary care physicians in the “paper without incentive” group, 10 in the “paper with incentive” group, 7 in the “paper or web-based with tablet” group, and 7 in the “paper or web-based via QR code” group.

XSL•FO
RenderX

Figure 1. Flowchart of the study.



The study sample consisted of 974 consecutive nonurgent patients who were recruited from these 39 medical practices, representing an average of 25 patients per practice (min=16, max=37). There were 251 patients in the "paper with incentive" group, 368 in the "paper without incentive" group, 187 in the "mixed mode with tablet" group, and 168 in the "mixed mode with QR code" group (Figure 1). Of these patients, 822 agreed to participate in the study, resulting in a participation rate of 84.4% (woman: 536/821, 65.3%; median age of 52, IQR 37-68; min-max=20-93; aged less than 40 years=239/819, 29.2%; 40-60 years=267/819, 32.6%; and more than 60 years=313/819, 38.2%). The distribution of women and men and by age group was similar in all 4 groups (*P* value=0.91 for sex and 0.08 for age group). The majority of participants were retirees (264/822; 32.1%), employees (227/822, 27.6%), and executives or intellectual professionals (166/822, 20.2%). Smaller groups included those not currently active in the workforce (68/822, 8.3%), intermediate professions (41/822, 5%), workers (28/822, 3.4%), artisans or business owners (25/822, 3%), and farmers (3/822; 0.4%).

The main results of the study are presented in Tables 1 and 2. Table 1 shows the differences between the combined paper group, the "mixed mode with tablet" group, and the "mixed mode with QR code" group. The data are organized first by the initial randomization group and then by the chosen questionnaire format. Table 2 focuses on comparing the "paper with incentive" group to the "paper without incentive" group.

XSL•FO
RenderX

Table 1. Participation and completion rates and median number of questions answered according to initial randomization group, chosen questionnaire format, sex, and age group[a].

	Participation rate, n/N (%)	Crude OR[b] (95% CI)	*P* value[c]	Number of patients having fully completed the questionnaire, n/N (%)	Median number of questions answered (IQR; min-max)	Crude OR (95% CI)	*P* value[d]	Adjusted OR (95% CI)	*P* value[e]
Initial randomization group (n=822)			.23				<.001		<.001
Paper (n=517)	517/619 (83.5)	Reference		516/517 (99.8)	48 (48-48; 47-48)	Reference		Reference	
Paper or tablet (n=156)	156/187 (83.4)	0.99 (0.55-1.80)		151/156 (96.8)	48 (48-48; 10-48)	0.05 (0-0.48)		0.06 (0-0.53)	
Paper or QR-code (n=149)	149/168 (88.7)	1.55 (0.91-2.63)		139/149 (93.3)	48 (48-48; 21-48)	0.02 (0-0.17)		0.03 (0-0.19)	
Chosen questionnaire format (n=822)							<.001		<.001
Paper (n=634)	N/A[f]	N/A	N/A	633/634 (99.8)	48 (48-48; 47-48)	Reference		Reference	
Web-based with tablet (n=105)	N/A	N/A	N/A	100/105 (95.2)	48 (48-48; 10-48)	0.03 (0-0.30)		0.03 (0-0.27)	
Web-based with QR code (n=83)	N/A	N/A	N/A	73/83 (88)	48 (48-48; 21-48)	0.01 (0-0.06)		0.01 (0-0.07)	
Sex (n=821)							.20		.17
Female (n=536)	N/A	N/A	N/A	529/536 (98.7)	48 (48-48; 10-48)	Reference		Reference	
Male (n=285)	N/A	N/A	N/A	277/285 (97.2)	48 (48-48; 21-48)	0.53 (0.17-1.62)		0.46 (0.14-1.47)	
Age group (years, n=819)							.06		.16
Less than 40 (n=239)	N/A	N/A	N/A	233/239 (97.5)	48 (48-48; 10-48)	Reference		Reference	
40-60 (n=267)	N/A	N/A	N/A	265/267 (99.3)	48 (48-48; 38-48)	5.06 (1.03-48.62)		3.99 (0.79-39.17)	
More than 60 (n=313)	N/A	N/A	N/A	308/313 (98.4)	48 (48-48; 21-48	2.35 (0.69-9.08)		1.92 (0.55-7.64)	

[a]The questionnaire consisted of 48 questions: 4 sociodemographic questions, the Sense of Coherence Scale (=17 questions), the Meat Attachment Questionnaire (=17 questions), and an intentionality questionnaire developed by our research team (=10 questions).

[b]OR: odds ratio.

[c]Univariable logistic regression (adjusted for intracluster correlation within practices).

[d]Univariable exact logistic regression (adjusted for intracluster correlation within practices).

[e]Multivariable exact logistic regression (adjusted for sex, age group, occupation, and intracluster correlation within practices).

[f]Participation rates are marked as "N/A" for the chosen questionnaire format, sex, and age group because participants who refused to participate did not indicate their preferred format, sex, or age. Therefore, participation rates, crude odds ratio, and *P* values by these characteristics could not be calculated.

XSL•FO
RenderX

Table 2. Comparison of participation and completion rates and median number of questions answered in the "paper with incentives" and "paper without incentives" groups[a].

	Participation rate, n/N (%)	Crude OR[b] (95% CI)	*P* value[c]	Number of patients having fully completed the questionnaire, n/N (%)	Median number of questions answered (IQR; min-max)	Crude OR (95% CI)	*P* value[d]
Paper group (n=517)	N/A[e]	N/A	.18	N/A	N/A	N/A	≥.99
Without incentives (n=300)	300/368 (81.5)	Reference		299/300 (99.7)	48 (48-48; 47-48)	Reference	
With incentives (n=217)	217/251 (86.5)	1.45 (0.84–2.49)		217/217 (100)	48 (48-48; 48-48)	1.74 (0-67.75)	

[a]The questionnaire consisted of 48 questions: 4 sociodemographic questions, the Sense of Coherence Scale (=17 questions), the Meat Attachment Questionnaire (=17 questions), and an intentionality questionnaire developed by our research team (=10 questions).

[b]OR: odds ratio.

[c]Univariable logistic regression (adjusted for intracluster correlation within practices).

[d]Univariable exact logistic regression (adjusted for intracluster correlation within practices).

[e]N/A: not applicable.

The participation rate ranged from 81.5% in the "paper without incentive" group to 88.7% in the "mixed mode with QR code" group, but the differences between the groups were not statistically significant. The incentive did not lead to a substantial increase in the participation rate. In this study, 51 patients (32.7%) in the "paper or web-based with tablet" group and 66 patients (44.3%) in the "paper or web-based via QR code" group preferred to complete the questionnaire using the paper version.

Overall, 806 patients (ie, 98.1% of patients) answered all 48 questions on the questionnaire. For the remaining 16 patients, the number of missing data ranged from 1 to 38. These patients were divided as follows: 1 patient belonged to the "paper without incentive" group, 5 to the "tablet" group, and 10 to the "QR code" group. The differences between groups based on the initial randomization and the questionnaire format chosen were small but statistically significant in both univariable and multivariable analyses. By contrast, the differences were not statistically significant for incentive, sex, and age group, but the number of patients with missing data was low.

Discussion

Main Findings

This study involved French primary care patients, recruited by research assistants and invited to complete a 48-question survey in the waiting room. We found that the participation rate was over 80% for all groups and that, when given the choice, a number of patients (n=117) preferred to complete the paper rather than the web-based questionnaire. We also found that the completion rate was 99.8% for the paper questionnaire and only slightly less for the web-based questionnaire. Finally, we found that the incentive had no influence on the participation rate or the number of questions answered.

Comparison With Existing Literature

Compared with other studies [12,23,24], we achieved excellent participation and completion rates in all groups (for participation rates, much higher than the 60% recommended in some publications [25]), despite a relatively lengthy self-administered questionnaire comprising 48 questions on a sensitive topic. The results of our study tend to reinforce the already known information that, among primary care patients, the response rate may not be affected by the length of the questionnaire, provided that the total duration is less than 15-20 minutes [26,27], which was the case in our study.

We found that the use of mixed options (paper or web-based), although intended to adapt more precisely to the preferences of each participant, did not appear to add value in terms of participation and completion rates. However, the web-based format (or mixed options) may be preferred for other reasons such as improved feasibility. Compared to paper questionnaires, the web-based format generally allows data to be obtained more cheaply (less printing, mailing, or typing costs), more quickly (real-time data tracking and less typing), and more accurately (the structured format minimizes incorrect entries and automatic data transfer minimizes data entry errors) [28]. Alongside the usual methods for collecting web-based data, QR codes have recently come into use in questionnaire surveys. They are simple and effective tools that are known to increase user engagement [29]. In a UK population-based maternity study, the inclusion of QR codes in the survey tended to lead to an increase in response rate, but this effect was limited and was probably also related to other factors (prior notification, short questionnaire, personalized study material, and additional reminder) [8]. In this English study, the initial response rate (30%) was considerably lower than ours.

The results of our study suggest that face-to-face recruitment with a research assistant achieves high participation and completion rates, providing support to participants. Previous studies have already highlighted that this strategy was associated with significantly higher participation rates [7,13,30]. Administering a face-to-face survey in the waiting room has the potential advantage of capturing patients' experiences when they are most accessible, with fewer competing demands than in other situations.

We found that, in the presence of a research assistant, neither the mixed mode, as discussed earlier, nor the unconditional nonmonetary incentive seemed to have any effect. Nonmonetary

XSL•FO
RenderX

incentives are recognized as important factors in improving the recruitment of participants to health-related surveys, with the impact being greater when the incentive is unconditional [31]. In our study, the high participation rate probably prevented us from observing a significant effect in the different methods tested. The presence of a research assistant seems to be the most powerful incentive, probably due to a desirability bias. This could be verified by carrying out the study using the incentives in the absence of a research assistant.

Significance and Recommendations for Future Research

This study makes a valuable contribution to the methodology of survey-based research in health care by underscoring the impact of recruitment strategies on both participation and completion rates. The notably high rates we observed across all groups suggest that face-to-face recruitment by a research assistant may be a critical factor for ensuring representativeness, especially in primary care settings where patient engagement can be challenging. Researchers conducting survey studies in similar environments may benefit from prioritizing in-person recruitment with trained personnel, as our findings indicate that this approach effectively mitigates common barriers to participation, such as lack of time. Additionally, while mixed mode survey options (including QR codes) did not significantly increase participation in this study, these tools still offer logistical advantages, such as ease of data collection, reduced error rates, and potential cost savings. For research teams looking to optimize both response rates and operational efficiency, mixed mode strategies may still be worthwhile. Future studies may benefit from investigating the impact of assistant-supported recruitment in different contexts or exploring alternative incentives, such as monetary incentives, that could further enhance response rates, particularly in larger or more geographically diverse samples.

Limitations

First, the results of our study cannot be generalized in all questionnaire-based surveys. Several studies showed that surveys carried out by recruiting participants by email, post, or telephone often led to mediocre participation rates for the web-based format [1-4] and mediocre completion rates for the paper format [1,2]. In this type of study, mixed mode options and incentives could be useful for positively influencing participation and completion rates. Second, selection bias is always possible in this type of survey, but in our opinion, it was reduced to a minimum by selecting physicians at random, including patients consecutively, and obtaining a high participation rate. In the context of our study, where patients were present in the waiting room, the selection biases usually described for electronic surveys in general practice do not apply [32]. Third, the study was carried out in only 1 region of France and only with primary care patients. The results obtained could have been different for other regions or countries, or for other study populations (physicians or patients visiting specialists). Finally, the use of simple randomization resulted in an unequal allocation of clusters across the study groups. While this does not affect the validity of randomization, it may have slightly reduced the power to detect differences between groups.

Conclusions

In conclusion, by recruiting patients in the waiting room with the help of a research assistant providing support to participants, we obtained a participation rate of over 80% in all groups. Neither the choice of response mode (ie, single or mixed, with or without a QR code) nor the use of incentives markedly influenced participation or completion rates. Future research would be useful to compare these rates with and without the presence of a research assistant in the waiting room to answer questions from participants. In terms of feasibility, the use of mixed options, including innovative methods such as QR codes, to offer participants the opportunity to choose, might be a relevant strategy, despite the results of our study showing no superiority in terms of participation and completion rates.

Acknowledgments

The authors of the study would like to sincerely thank all the primary care physicians and patients who collaborated on this project. The authors give special thanks to Amir Moussa for his involvement in this project. Patients or the public were not involved in the design, conduct, reporting, or dissemination plans of our research.

Data Availability

The data associated with this article are available from the corresponding author upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire used in the study.
[DOCX File , 147 KB - ijmr_v14i1e67981_app1.docx]

References

XSL•FO
RenderX

1. Sebo P, Maisonneuve H, Cerutti B, Fournier JP, Senn N, Haller DM. Rates, delays, and completeness of general practitioners' responses to a postal versus web-based survey: a randomized trial. J Med Internet Res 2017;19(3):e83 [FREE Full text] [doi: 10.2196/jmir.6308] [Medline: 28330830]
2. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of internet versus paper-and-pencil versions. J Med Internet Res 2007;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
3. Bergeson SC, Gray J, Ehrmantraut LA, Laibson T, Hays RD. Comparing web-based with mail survey administration of the Consumer Assessment of Healthcare Providers and Systems (CAHPS) Clinician and Group survey. Prim Health Care 2013;3(01) [FREE Full text] [doi: 10.4172/2167-1079.1000132] [Medline: 24078901]
4. McMahon SR, Iwamoto M, Massoudi MS, Yusuf HR, Stevenson JM, David F, et al. Comparison of e-mail, fax, and postal surveys of pediatricians. Pediatrics 2003;111:e299-e303. [doi: 10.1542/peds.111.4.e299] [Medline: 12671142]
5. Brtnikova M, Crane LA, Allison MA, Hurley LP, Beaty BL, Kempe A. A method for achieving high response rates in national surveys of U.S. primary care physicians. PLoS One 2018;13(8):e0202755 [FREE Full text] [doi: 10.1371/journal.pone.0202755] [Medline: 30138406]
6. Brehaut JC, Graham ID, Visentin L, Stiell IG. Print format and sender recognition were related to survey completion rate. J Clin Epidemiol 2006;59(6):635-641. [doi: 10.1016/j.jclinepi.2005.04.012] [Medline: 16713527]
7. Sitzia J, Wood N. Response rate in patient satisfaction research: an analysis of 210 published studies. Int J Qual Health Care 1998;10(4):311-317. [doi: 10.1093/intqhc/10.4.311] [Medline: 9835247]
8. Harrison S, Henderson J, Alderdice F, Quigley MA. Methods to increase response rates to a population-based maternity survey: a comparison of two pilot studies. BMC Med Res Methodol 2019;19(1):65 [FREE Full text] [doi: 10.1186/s12874-019-0702-3] [Medline: 30894130]
9. Booker QS, Austin JO, Balasubramanian BA. Survey strategies to increase participant response rates in primary care research studies. Fam Pract 2021;38(5):699-702. [doi: 10.1093/fampra/cmab070] [Medline: 34213547]
10. Parker MJ, Manan A, Urbanski S. Prospective evaluation of direct approach with a tablet device as a strategy to enhance survey study participant response rate. BMC Res Notes 2012;5(1):605 [FREE Full text] [doi: 10.1186/1756-0500-5-605] [Medline: 23110849]
11. Markwardt HS, Taghavi SE, Williams AP, Olivares MN, McDuffee PR, Hall BC. Tablet versus quick response code: Difference in research participation rates? J Technol Behav Sci 2023;8(2):178-181 [FREE Full text] [doi: 10.1007/s41347-023-00311-8] [Medline: 37215394]
12. Kelly BJ, Fraze TK, Hornik RC. Response rates to a mailed survey of a representative sample of cancer patients randomly drawn from the Pennsylvania Cancer Registry: a randomized trial of incentive and length effects. BMC Med Res Methodol 2010;10:65 [FREE Full text] [doi: 10.1186/1471-2288-10-65] [Medline: 20630101]
13. Anhang Price R, Quigley DD, Hargraves JL, Sorra J, Becerra-Ornelas AU, Hays RD, et al. A systematic review of strategies to enhance response rates and representativeness of patient experience surveys. Med Care 2022;60(12):910-918 [FREE Full text] [doi: 10.1097/MLR.0000000000001784] [Medline: 36260705]
14. Perez-Alba E, Nuzzolo-Shihadeh L, Espinosa-Mora JE, Camacho-Ortiz A. Use of self-administered surveys through QR code and same center telemedicine in a walk-in clinic in the era of COVID-19. J Am Med Inform Assoc 2020;27(6):985-986 [FREE Full text] [doi: 10.1093/jamia/ocaa054] [Medline: 32282922]
15. Turkistani KA. The use of innovative cost-saving audience response system in orthodontic case-based learning: a potential approach in distance learning. J Orthod Sci 2021;10:4 [FREE Full text] [doi: 10.4103/jos.JOS_44_20] [Medline: 34084760]
16. Masih EA. Nurse Educ Today 2022;111:105305 [FREE Full text] [doi: 10.1016/j.nedt.2022.105305] [Medline: 35220149]
17. Lacroix K, Gifford R. Reducing meat consumption: identifying group-specific inhibitors using latent profile analysis. Appetite 2019;138:233-241. [doi: 10.1016/j.appet.2019.04.002] [Medline: 30965045]
18. Prokop-Dorner A, Piłat-Kobla A, Zając J, Luśtyk M, Valli C, Łapczuk A, et al. Values and preferences influencing willingness to change red and processed meat consumption in response to evidence-based information: a mixed methods study. Public Health Nutr 2022;25(8):2084-2098. [doi: 10.1017/s1368980022000866]
19. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
20. Kirkwood BR, Sterne JAC, Kirkwood BR. Essential Medical Statistics. Hoboken, NJ: Blackwell Science; 2003.
21. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. J Clin Epidemiol 2004;57(8):785-794. [doi: 10.1016/j.jclinepi.2003.12.013] [Medline: 15485730]
22. Mehta CR, Patel NR. Exact logistic regression: Theory and examples. Stat Med 1995;14(19):2143-2160. [doi: 10.1002/sim.4780141908] [Medline: 8552893]
23. Kazzazi F, Haggie R, Forouhi P, Kazzazi N, Malata C. Utilizing the total design method in medicine: maximizing response rates in long, non-incentivized, personal questionnaire postal surveys. Patient Relat Outcome Meas 2018;Volume 9:169-172. [doi: 10.2147/prom.s156109]
24. Haller DM, Sanci LA, Patton GC, Sawyer SM. Text message communication in primary care research: a randomized controlled trial. Fam Pract 2009;26(4):325-330. [doi: 10.1093/fampra/cmp040] [Medline: 19542193]

XSL•FO
RenderX

25. Fincham JE. Response rates and responsiveness for surveys, standards, and the journal. Am J Pharm Educ 2008;72(2):43 [FREE Full text] [doi: 10.5688/aj720243] [Medline: 18483608]
26. Beebe TJ, Rey E, Ziegenfuss JY, Jenkins S, Lackore K, Talley NJ, et al. Shortening a survey and using alternative forms of prenotification: impact on response rate and quality. BMC Med Res Methodol 2010;10(1):50 [FREE Full text] [doi: 10.1186/1471-2288-10-50] [Medline: 20529365]
27. Guo Y, Kopec JA, Cibere J, Li LC, Goldsmith CH. Population survey features and response rates: a randomized experiment. Am J Public Health 2016;106(8):1422-1426. [doi: 10.2105/ajph.2016.303198]
28. Determann D, Lambooij MS, Steyerberg EW, de Bekker-Grob EW, de Wit GA. Impact of survey administration mode on the results of a health-related discrete choice experiment: online and paper comparison. Value Health 2017;20(7):953-960 [FREE Full text] [doi: 10.1016/j.jval.2017.02.007] [Medline: 28712625]
29. Cata T, Patel PS, Sakaguchi T. QR code: A new opportunity for effective mobile marketing. J Mobile Technol Knowl Soc 2013:ID748267 [FREE Full text] [doi: 10.5171/2013.748267]
30. Heerman WJ, Jackson N, Roumie CL, Harris PA, Rosenbloom ST, Pulley J, et al. Recruitment methods for survey research: Findings from the mid-south clinical data research network. Contemp Clin Trials 2017;62:50-55. [doi: 10.1016/j.cct.2017.08.006] [Medline: 28823925]
31. Zheng G, Oksuzyan S, Hsu S, Cloud J, Jewell MP, Shah N, et al. Self-reported interest to participate in a health survey if different amounts of cash or non-monetary incentive types were offered. J Urban Health 2018;95(6):837-849 [FREE Full text] [doi: 10.1007/s11524-018-0237-7] [Medline: 29654397]
32. Slater M, Kiran T. Measuring the patient experience in primary care: comparing e-mail and waiting room survey delivery in a family health team. Can Fam Physician 2016;62(12):e740-e748 [FREE Full text] [Medline: 27965350]

Edited by T de Azevedo Cardoso; submitted 25.10.24; peer-reviewed by A Dabbagh, M Sandri, SY Choi; comments to author 04.12.24; revised version received 26.12.24; accepted 17.01.25; published 25.02.25.

Please cite as:
Sebo P, Tudrej B, Bernard A, Delaunay B, Dupuy A, Malavergne C, Maisonneuve H
Increasing Participation and Completion Rates in Questionnaire Surveys of Primary Care Patients: Cluster-Randomized Study
Interact J Med Res 2025;14:e67981
URL: https://www.i-jmr.org/2025/1/e67981
doi:10.2196/67981
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/), 25.02.2025. 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 of Modifiable Lifestyle and Metabolic Factors With the Risk of Developing Sepsis: 2-Sample Mendelian Randomized Study

Haifeng Lv[1*]; Jing Liu[2*]; Yelin Cao[1*]; Weina Fan[1]; Guojie Shen[1]; Feifei Wang[1]; Qingqing Ye[1]; Xiaoliang Wu[1]; Kaijin Xu[3,4]

[1]Department of Intensive Care Unit, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

[2]Department of Hepatology, The Affiliated Hospital of Hangzhou Normal University, Hangzhou, China

[3]State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, China-Singapore Belt and Road Joint Laboratory on Infection Research and Drug Development, National Medical Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, No.79 Qingchun Road, Shangcheng District, Zhejiang Province, Hangzhou, China

[4]Yuhang Institute for Collaborative Innovation and Translational Research in Life Sciences and Technology, Hangzhou, China

[*]these authors contributed equally

Corresponding Author:
Kaijin Xu
State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Diseases, China-Singapore Belt and Road Joint Laboratory on Infection Research and Drug Development, National Medical Center for Infectious Diseases, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, The First Affiliated Hospital, Zhejiang University School of Medicine, No.79 Qingchun Road, Shangcheng District, Zhejiang Province, Hangzhou, China

Abstract

Background: Sepsis is a life-threatening condition characterized by organ dysfunction resulting from dysregulated host response to infections. Approximately 48.9 million people worldwide are diagnosed with sepsis annually, leading to 11 million deaths and representing 19.7% of all global deaths. No specific, effective treatments for sepsis, which has a poor prognosis, are available.

Objective: The study aimed to systematically explore the association between genetically predicted modifiable risk factors and sepsis.

Methods: Univariable 2-sample Mendelian randomization (MR) analysis was performed to explore the association between 30 modifiable risk factors (12 lifestyle, 3 educational and psychological, and 15 metabolic factors) and sepsis. Heterogeneity was evaluated using the Cochran *Q* analysis. Sensitivity analyses were conducted using the MR-Egger regression intercept tests and leave-one-out analyses. Additionally, multivariable MR analyses were performed to adjust for genetic associations between the instruments and obesity.

Results: Genetically predicted smoking (odds ratio [OR] 1.20, 95% CI 1.06 - 1.36; *P*=.005), a higher number of cigarettes smoked daily (OR 1.70, 95% CI 1.29 - 2.23; *P*<.001), a higher overall health rating (OR 2.19, 95% CI 1.61 - 2.98; *P*<.001), BMI (OR 1.50, 95% CI 1.38 - 1.63; *P*<.001), waist circumference (OR 1.70, 95% CI 1.53 - 1.89; *P*<.001), whole body fat mass (OR 1.50, 95% CI 1.37 - 1.64; *P*<.001), trunk fat mass (OR 1.48, 95% CI 1.36 - 1.62; *P*<.001), arm fat mass (OR 1.57, 95% CI 1.43 - 1.71; *P*<.001), and leg fat mass (OR 1.69, 95% CI 1.51 - 1.90; *P*<.001) were associated with increased sepsis risk. However, light physical activity (OR 0.26, 95% CI 0.08 - 0.83; *P*=.03), higher education attainment (OR 0.52, 95% CI 0.40 - 0.67; *P*<.001), and high-density lipoprotein cholesterol (OR 0.91, 95% CI 0.84 - 0.98; *P*=.02) exhibited protective effects against sepsis. Using a multivariate analysis of obesity traits, the waist circumference (OR 2.16, 95% CI 1.18 - 3.96; *P*=.01) was an independent risk factor of sepsis.

Conclusions: Our study demonstrated that genetic predictors of lifestyle (smoking and physical activity), educational level, and metabolic factors (waist circumference and high-density lipoprotein cholesterol) exhibited a causal association with sepsis risk. Future research should further investigate the underlying mechanisms of these associations to inform more effective preventive strategies against sepsis.

(Interact J Med Res 2025;14:e72244) doi:10.2196/72244

KEYWORDS

Mendelian randomization; sepsis; modifiable factors; lifestyle; metabolism

XSL•FO
RenderX

Introduction

Sepsis is a life-threatening condition characterized by organ dysfunction resulting from a dysregulated host response to infections [1]. Epidemiological studies indicate that approximately 48.9 million people worldwide were diagnosed with sepsis annually, leading to 11 million deaths and representing 19.7% of all global deaths [2]. In China, sepsis affects 20% of patients in intensive care units, with a 90-day mortality rate of 35.5% [3]. The substantial epidemiological burden underscores the critical importance of sepsis as a public health issue [4]. Despite decades of research, current therapies remain limited to organ support and antimicrobial interventions, without any targeted treatment available [5-7]. These limited therapeutic options highlight the urgent need for shifting focus toward modifiable risk factors for prevention. The World Health Organization 2020 Sepsis Report specifically highlights vulnerable populations and health care seekers, while outlining future research directions and priorities in sepsis epidemiology. Although sepsis manifests diversely and can be life-threatening, it remains preventable and reversible [4].

Host risk factors for sepsis are multifaceted, encompassing extremes of age, immunosuppression, and comorbidities [8,9]. Emerging evidence suggests that socioeconomic determinants (eg, education level and occupation) and metabolic disorders (eg, obesity and diabetes) may influence susceptibility to sepsis [10-16]. However, the existing observational studies report inconsistent conclusions. For instance, although lower socioeconomic status is broadly linked to higher sepsis incidence [10,11], another study paradoxically found no association using adjusted models [17]. Metabolic comorbidities further illustrate this ambiguity. Although obesity is widely implicated in sepsis risk [18], some observational studies found that individuals with obesity showed a lower incidence and fatality than did those with underweight among patients with sepsis. These discrepancies may stem from methodological limitations in observational research, such as residual confounding and reverse causality. For instance, sepsis-induced systemic inflammation could transiently alter metabolic markers, obscuring true causal directions.

The core idea of Mendelian randomization (MR) is to use genetic variants (typically single-nucleotide polymorphisms [SNPs]) as instrumental variables to infer whether a potential causal association exists between an exposure and a health outcome. Recent MR studies have explored individual factors such as obesity, diabetes mellitus, and lipid profiles in patients with sepsis [19-23]; however, critical gaps persist. First, lifestyle factors (eg, smoking and diet) remain understudied despite their clinical relevance. Second, the existing MR analyses of sepsis predominantly examine isolated risk factors, neglecting potential synergies between socioeconomic, metabolic, and behavioral determinants [8,9,24-28]. Third, the interaction between modifiable factors and genetic predisposition remains unexplored, limiting personalized preventive strategies.

To address these gaps, we conducted a comprehensive 2-sample MR analysis of 30 modifiable factors spanning lifestyle, socioeconomic status, and metabolic health. This approach not only circumvents confounding biases but also systematically evaluates multifactorial contributions to sepsis pathogenesis. Our findings aimed to clarify controversial associations in prior literature and prioritize actionable targets for intervention.

Methods

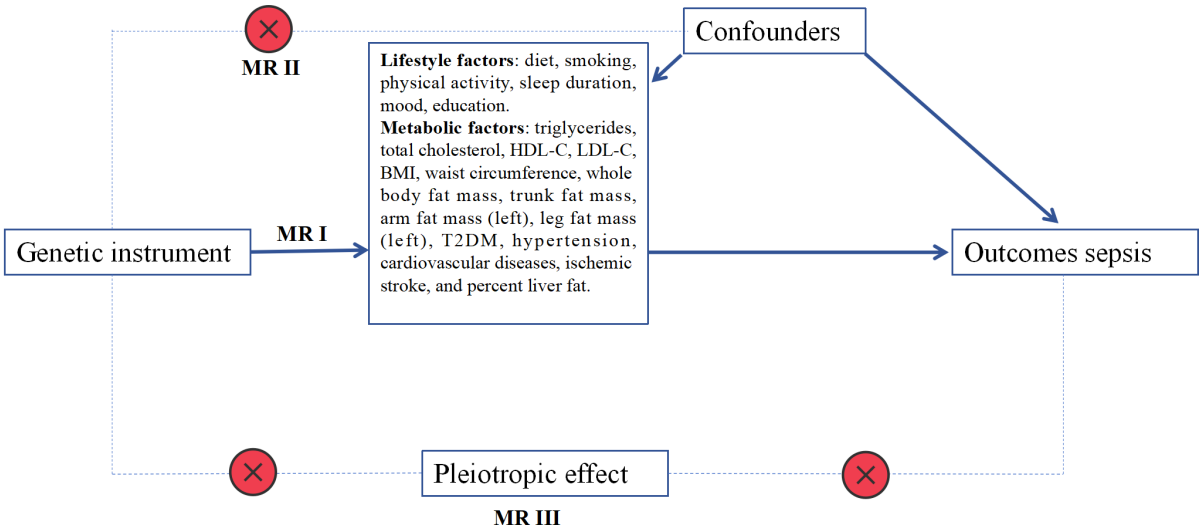
MR Design

MR is based on three key assumptions: genetic variants are (1) linked to risk factors, (2) independent of confounding factors, and (3) influence outcomes solely through risk factors (Figure 1). A total of 30 prominent modifiable risk factors were included and grouped into lifestyle and metabolic factors. This study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization (STROBE-MR) guidelines. Initially, we screened instrumental variables for MR analysis using lifestyle factors and metabolic comorbidities as “exposures” and sepsis as “outcomes.” Subsequently, we conducted a reverse MR analysis, with sepsis as the “exposure” and lifestyle and metabolic factors as the “outcomes.”

XSL•FO
RenderX

Figure 1. Overview of the design and methods used in this Mendelian randomization (MR) study. MR analysis was used to explore the causal relationships, including the following groups: lifestyles and metabolic factors. Solid arrows represent causal effects, and dashed arrows and crosses represent causal effects prohibited by MR assumptions II and III. HDL-C: high-density lipoprotein cholesterol; LDL-C: low-density lipoprotein cholesterol; T2DM: type 2 diabetes mellitus.

Figure 1



Data Sources

Genome-wide association study (GWAS) data of sepsis were obtained from the OpenGWAS database, a dataset published by the UK Biobank in 2021, comprising a sample of 486,484 participants from European populations. Additionally, we used 30 parameters from the Integrative Epidemiology Unit OpenGWAS Project as our lifestyle, education and psychology, and metabolic traits. Data are listed in Table 1.

XSL•FO
RenderX

Table . Characteristics of the genome - wide association study (GWAS) summary data in this study.

Exposure	Ethnicity	GWAS ID	Total population, n	SNPs[a], n
Diet				
Alcohol consumption	European	ieu-b-73	335,394	11,887,865
Coffee consumption	European	ukb-b-5237	428,860	9,851,867
Milk consumption	South Asian	ukb-e-100520_CSA	1469	9,797,409
Sweets consumption	African American	ukb-e-102330_AFR	1207	15,533,528
Smoking				
Smoking initiation	European	ieu-b-4877	607,291	11,802,365
Age of smoking	European	ieu-b-24	341,427	11,894,779
Number of cigarettes daily	European	ukb-b-6019	108,946	9,851,867
Physical activity				
Light activity	European	ukb-b-11495	460,376	9,851,867
Moderate to vigorous	European	ebi-a-GCST006097	377,234	11,808,007
Sleep habit				
Insomnia	European	ukb-b-3957	462,341	9,851,867
Nap during day	European	ukb-b-4616	462,400	9,851,867
Sleep duration	European	ukb-b-4424	460,099	9,851,867
Physical and mental				
Unhealthy emotions	European	ukb-b-6991	459,560	9,851,867
Overall health rating	European	ukb-b-16489	458,079	9,851,867
College or university	European	ukb-b-6306	460,844	9,851,867
Blood lipid traits				
Triglycerides	European	ieu-b-4850	78,700	7,892,037
Total cholesterol	European	met-d-Total_C	115,078	12,321,875
HDL - c[b]	European	met-d-HDL_C	115,078	12,321,875
LDL - c[c]	European	ie,u-b-110	440,546	12,321,875
Obesity traits				
BMI	European	ukb-b-19953	461,460	9,851,867
Waist circumference	European	ukb-b-9405	462,166	9,851,867
Whole body fat mass	European	ukb-b-19393	454,137	9,851,867
Trunk fat mass	European	ukb-b-20044	454,588	9,851,867
Arm fat mass (left)	European	ukb-b-8338	454,684	9,851,867
Leg fat mass (left)	European	ukb-b-7212	454,823	9,851,867
Metabolic syndrome				
T2DM[d]	European	ebi-a-GCST010118	433,540	11,222,507
Hypertension	European	finn-b-I9_HYPTENS	162,837	16,380,466
Cardiovascular diseases	European	finn-b-I9_CVD	107,684	16,380,466
Ischemic stroke	European	ebi-a-GCST006908	440,328	8,296,492
Percent liver fat	European	ebi-a-GCST90016673	32,858	9,275,407

[a]SNP: single-nucleotide polymorphism.

[b]HDL - C: high-density lipoprotein cholesterol.

[c]LDL - C: low-density lipoprotein cholesterol.

XSL•FO
RenderX

[d]T2DM: type 2 diabetes mellitus.

Selection of Genetic Variants

In this MR analysis, instrumental variables were used to investigate the association between potentially modifiable risk factors and sepsis. These risk traits were categorized into three main parts: (1) lifestyle factors encompassing 4 dietary traits, 3 smoking-related traits, 2 pertaining to physical activity, and 3 regarding sleep habits; (2) educational and psychological factors, including physical and mental well-being and education level; and (3) metabolic factors comprising 4 traits linked to blood lipid parameters, 6 associated with obesity traits, and 5 related to metabolic comorbidities. SNPs were selected as independent genetic predictors based on the following criteria: first, a genome-wide significance threshold of $P<5\times10^{-8}$ was applied. If there were insufficient significant SNPs under this threshold, a threshold of $P<1\times10^{-5}$ was used. Second, the clump function was used to test for linkage disequilibrium with a threshold of $R^2<0.001$ and a distance of 10,000 kb. Third, SNPs related to outcomes were excluded using the linkage disequilibrium link database to avoid confounding factors. Fourth, SNPs with an *F* statistic <10 were excluded to avoid bias from weak instruments. The proportion of exposure variance explained by genetic instruments (R^2) was calculated to quantify the strength of the genetic tools. The *F* statistic for each SNP was calculated to assess the strength of the selected instruments.

Ethical Considerations

This study is a secondary analysis of publicly available GWAS summary statistics obtained from the Integrative Epidemiology Unit OpenGWAS database [29] as shown in Table 1. These datasets consist of deidentified data from participant studies approved by an ethics committee concerning human experimentation, comply with the ethical principles of the Declaration of Helsinki, and were approved by the Ethics Committee of the First Affiliated Hospital of Zhejiang University (2025B-1097).

Statistical Analysis

In this study, all analyses were performed using the *TwoSampleMR* package in R (version 4.2.1; R Foundation for Statistical Computing). The random-effects inverse variance weighted (IVW) method was used as the primary outcome; this method assumed that all instrumental variables (SNPs) were valid instruments (ie, satisfying the exclusion restriction assumption) and estimated the causal effect between exposure and outcome by calculating the weighted regression slope of the SNP-outcome associations on the SNP-exposure associations. Although the IVW approach offered the highest statistical power among MR methods, it was sensitive to directional pleiotropy, which might have biased causal estimates if invalid SNPs were present. Therefore, the MR-Egger and weighted median methods were used to refine the IVW estimation. These alternative methods offered more dependable estimates across a broader spectrum of scenarios, albeit with a trade-off of reduced efficiency, resulting in wider CIs. MR-Egger permitted all genetic variants to exhibit pleiotropy; however, pleiotropy must be independent of variant exposure associations. Meanwhile, the weighted median method allowed the incorporation of potentially invalid instrumental variables under the assumption that at least half of the instruments used in the MR analysis were valid. We performed reverse MR analyses by swapping the roles of the original exposure and outcome variables; IVW analysis tests were conducted using the same set of SNPs.

Heterogeneity was evaluated using the Cochran *Q* analysis. Sensitivity analyses were conducted using MR-Egger regression intercept tests and leave-one-out analyses, with statistical significance set at $P<.05$. Additionally, multivariable MR analyses were performed to adjust for the genetic association of the instruments with the BMI.

Results

Baseline Characteristics

We assessed 30 potentially modifiable risk factors to investigate their causal associations with sepsis and categorized them into 3 groups (Table 1): lifestyle, educational and psychological, and metabolic factors. The number of SNPs ranged between 7,892,037 and 15,533,528 (Table 1). Notably, the *F* statistics for all considered traits exceeded 10, indicating the absence of a potential weak instrumental bias.

Lifestyle Factors

Regarding the investigation of lifestyle factors, our findings revealed that genetically predicted smoking initiation (odds ratio [OR] 1.20, 95% CI 1.06 - 1.36; $P=.005$) and a higher number of cigarettes smoked daily (OR 1.70, 95% CI 1.29 - 2.23; $P<.001$) were associated with an increased risk of sepsis. Conversely, genetically predicted light physical activity (OR 0.26, 95% CI 0.08 - 0.83; $P=.02$) was associated with a decreased risk of sepsis. These results remained consistent across the random-effects IVW and weighted median analyses (Table 2). Notably, no significant causal association was detected between genetically predicted alcohol, coffee, milk, and sweet consumptions; age at smoking; moderate to vigorous physical activity; insomnia; daytime napping; sleep duration; and sepsis (all $P>.05$). We did not find statistically significant results when investigating reverse causality validation ($P>.05$).

XSL•FO
RenderX

Table . Mendelian randomization (MR) estimates for the causal effect of modifiable lifestyle and metabolic factors on sepsis.

Exposure	SNPs[a], n	IVW[b], OR[c] (95% CI)	*P* value	WM[d], OR (95% CI)	*P* value	MR-Egger, OR (95% CI)	*P* value	Cochran *Q*–derived *P* value	MR-Egger intercept–derived *P* value
Diet									
Alcohol consumption	35	1.00 (0.76 - 1.32)	.99	1.04 (0.69 - 1.57)	.85	1.02 (0.64 - 1.61)	.94	.34	.92
Coffee consumption	39	1.14 (0.79 - 1.64)	.48	1.10 (0.71 - 1.71)	.67	0.88 (0.42 - 1.83)	.73	.06	.43
Milk consumption	5	0.97 (0.92 - 1.03)	.29	0.95 (0.89 - 1.02)	.19	1.29 (0.73 - 2.28)	.45	.71	.40
Sweets consumption	11	0.99 (0.97 - 1.01)	.47	0.99 (0.97 - 1.02)	.64	1.05 (0.99 - 1.11)	.110	.25	.06
Smoking									
Smoking initiation	86	1.20 (1.06 - 1.36)	.005	1.22 (1.02 - 1.47)	.034	1.72 (0.90 - 3.28)	.10	.41	.27
Age of smoking	7	0.89 (0.46 - 1.71)	.723	1.25 (0.56 - 2.79)	.59	0.32 (0.04 - 2.68)	.34	.28	.37
Number of cigarettes daily	6	1.70 (1.29 - 2.23)	<.001	1.72 (1.25 - 2.36)	<.001	2.30 (1.31 - 4.04)	.04	.86	.29
Physical activity									
Light activity	12	0.26 (0.08 - 0.83)	.03	0.12 (0.03 - 0.58)	.007	1.12 (0.02 - 78.84)	.96	.74	.50
Moderate to vigorous	19	0.98 (0.61 - 1.57)	.92	1.10 (0.58 - 2.06)	.78	0.27 (0.03 - 2.58)	.27	.68	.27
Sleep habit									
Insomnia	38	1.08 (0.63 - 1.87)	.77	1.02 (0.50 - 2.07)	.95	0.52 (0.09 - 2.94)	.46	.02	.39
Nap during day	88	1.04 (0.74 - 1.45)	.84	0.93 (0.55 - 1.57)	.79	1.45 (0.46 - 4.55)	.53	.45	.55
Sleep duration	65	0.80 (0.57 - 1.11)	.19	0.83 (0.50 - 1.36)	.45	1.17 (0.32 - 4.32)	.81	.27	.55
Physical and mental									
Unhealthy emotions	41	1.36 (0.67 - 2.74)	.40	1.22 (0.45 - 3.30)	.70	0.48 (0.01 - 22.66)	.71	.81	.59
Overall health rating	106	2.19 (1.61 - 2.98)	<.001	1.63 (1.09 - 2.44)	.02	2.33 (0.49 - 11.00)	.29	.02	.94
College or university	242	0.52 (0.40 - 0.67)	<.001	0.55 (0.37 - 0.82)	.003	0.52 (0.18 - 1.51)	.23	.85	.97
Blood lipid traits									
Triglycerides	38	1.00 (0.89 - 1.13)	.94	1.02 (0.88 - 1.18)	.80	1.14 (0.91 - 1.42)	.25	.008	.20
Total cholesterol	359	0.97 (0.89 - 1.05)	.47	0.98 (0.87 - 1.10)	.70	1.00 (0.88 - 1.14)	.99	.02	.51
HDL - c[e]	91	0.91 (0.84 - 0.98)	.02	0.89 (0.79 - 0.99)	.04	0.86 (0.76 - 0.98)	.02	.01	.33
LDL - c[f]	173	1.05 (0.97 - 1.13)	.22	1.06 (0.93 - 1.20)	.36	1.08 (0.97 - 1.21)	.17	.39	.45
Obesity traits									
BMI	432	1.50 (1.38 - 1.63)	<.001	1.47 (1.25 - 1.72)	<.001	1.47 (1.17 - 1.85)	<.001	.21	.85

XSL•FO
RenderX

Exposure	SNPs[a], n	IVW[b], OR[c] (95% CI)	*P* value	WM[d], OR (95% CI)	*P* value	MR-Egger, OR (95% CI)	*P* value	Cochran *Q*–derived *P* value	MR-Egger intercept–derived *P* value
Waist circumference	357	1.70 (1.53 - 1.89)	<.001	1.64 (1.36 - 1.98)	<.001	1.52 (1.12 - 2.05)	.008	.32	.43
Whole body fat mass	412	1.50 (1.37 - 1.64)	<.001	1.51 (1.31 - 1.75)	<.001	1.50 (1.16 - 1.93)	.002	.05	.99
Trunk fat mass	399	1.48 (1.36 - 1.62)	<.001	1.51 (1.31 - 1.75)	<.001	1.02 (1.31 - 1.68)	.03	.08	.29
Arm fat mass (left)	402	1.57 (1.43 - 1.71)	<.001	1.52 (1.30 - 1.77)	<.001	1.57 (1.22 - 2.00)	<.001	.14	.99
Leg fat mass (left)	397	1.69 (1.51-1.90)	<.001	1.65 (1.37-1.99)	<.001	1.34 (1.83-2.53)	<.001	.11	.59
Metabolic syndrome									
T2DM[g]	148	1.03 (0.99 - 1.07)	.16	1.05 (0.98 - 1.11)	.14	1.07 (0.99 - 1.15)	.11	.11	.31
Hypertension	55	1.00 (0.94 - 1.06)	.98	0.97 (0.90 - 1.06)	.56	0.86 (0.70 - 1.05)	.15	.52	.13
Cardiovascular diseases	9	1.09 (0.91 - 1.32)	.35	0.96 (0.75 - 1.22)	.75	0.52 (0.21 - 1.32)	.21	.66	.16
Ischemic stroke	8	1.02 (0.88 - 1.18)	.78	0.99 (0.82 - 1.19)	.90	1.10 (0.37 - 3.24)	.87	.81	.90
Percent liver fat	10	1.02 (0.93 - 1.11)	.66	1.02 (0.91 - 1.13)	.76	0.96 (0.84 - 1.10)	.59	.56	.30

[a]SNP: single-nucleotide polymorphism.

[b]IVW: inverse variance weighted.

[c]OR: odds ratio.

[d]WM: weighted median.

[e]HDL - c: high - density lipoprotein cholesterol.

[f]LDL - c: low-density lipoprotein cholesterol.

[g]T2DM: type 2 diabetes mellitus.

Educational and Psychological Factors

Considering the investigation of educational and psychological factors, we found that genetically predicted higher overall health rating (OR 2.19, 95% CI 1.61 - 2.98; *P*<.001) was associated with an increased risk of sepsis (Table 2). Conversely, genetically predicted higher education attainment (individuals with a college or university degree; OR 0.52, 95% CI 0.40 - 0.67; *P*<.001) was associated with decreased risk of sepsis. Notably, no significant causal association was found between genetically driven unhealthy emotions and sepsis (*P*=.40). We did not find statistically significant results when investigating reverse causality validation (*P*>.05).

Metabolic Factors

We found an increased risk of sepsis associated with genetically predicted obesity. All fat mass–related traits were significantly associated with an increased risk of sepsis. The odds of developing sepsis increased with every 1 - SD increment in the BMI (OR 1.50, 95% CI 1.38 - 1.63; *P*<.001), waist circumference (OR 1.70, 95% CI 1.53 - 1.89; *P*<.001), whole body fat mass (OR 1.50, 95% CI 1.37 - 1.64; *P*<.001), trunk fat mass (OR 1.48, 95% CI 1.36 - 1.62; *P*<.001), arm fat mass (left; OR 1.57, 95% CI 1.43 - 1.71; *P*<.001), and leg fat mass (left; OR 1.69, 95% CI 1.51 - 1.90; *P*<.001; Table 2). However, high-density lipoprotein cholesterol (HDL-C; OR 0.91, 95% CI 0.84 - 0.98; *P*=.02) exhibited a protective effect against sepsis. In contrast, no significant causal association was observed between genetically predicted triglycerides, total cholesterol, low-density lipoprotein cholesterol, type 2 diabetes mellitus, hypertension, cardiovascular diseases, ischemic stroke, percent liver fat, and sepsis (all *P*>.05).

Reverse MR Analysis

To validate the reliability of the exposure-outcome causal direction, we performed reverse MR analysis by swapping the roles of the original exposure and outcome variables, and IVW analysis tests were conducted using the same set of SNPs. We did not find statistically significant results when investigating reverse causality validation (*P*>.05).

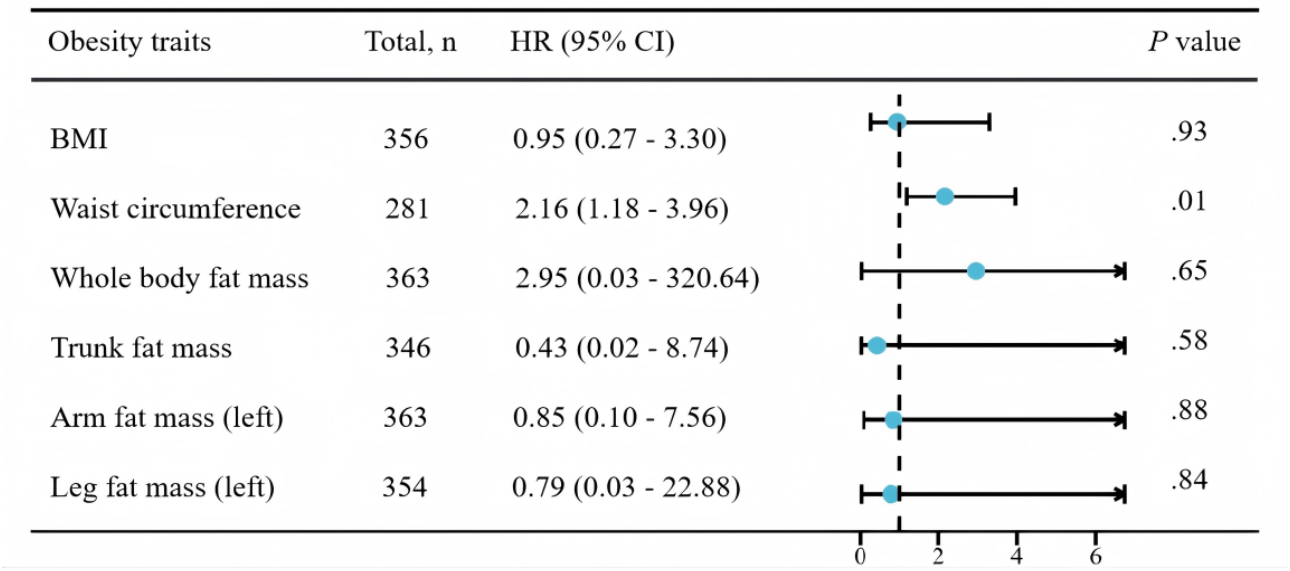
Multivariable MR Analysis

Considering the mutual influence of various obesity indicators, we conducted a multivariate MR analysis of BMI, waist

XSL•FO
RenderX

circumference, whole body fat mass, trunk fat mass, arm fat mass (left), and leg fat mass (left). The results revealed that the waist circumference (OR 2.16, 95% CI 1.18 - 3.96; *P*=.01) was an independent risk factor of sepsis (Figure 2).

Figure 2. The association between obesity traits (BMI, waist circumference, whole body fat mass, trunk fat mass, arm fat mass, and leg fat mass) and sepsis by multivariable Mendelian randomization. Odds ratios (ORs) represent the associations with sepsis.



Heterogeneity and Pleiotropy Test

Although some results displayed heterogeneity based on the Cochran *Q* test (*P*>.05), the use of random-effects IVW as the primary outcome rendered this heterogeneity acceptable and did not undermine the validity of the MR estimates in this study (Table 2). In addition, the results were subjected to pleiotropy testing, demonstrating that those with significance did not exhibit pleiotropy. Therefore, assuming the absence of horizontal pleiotropy in this study is reasonable.

Sensitivity Analysis

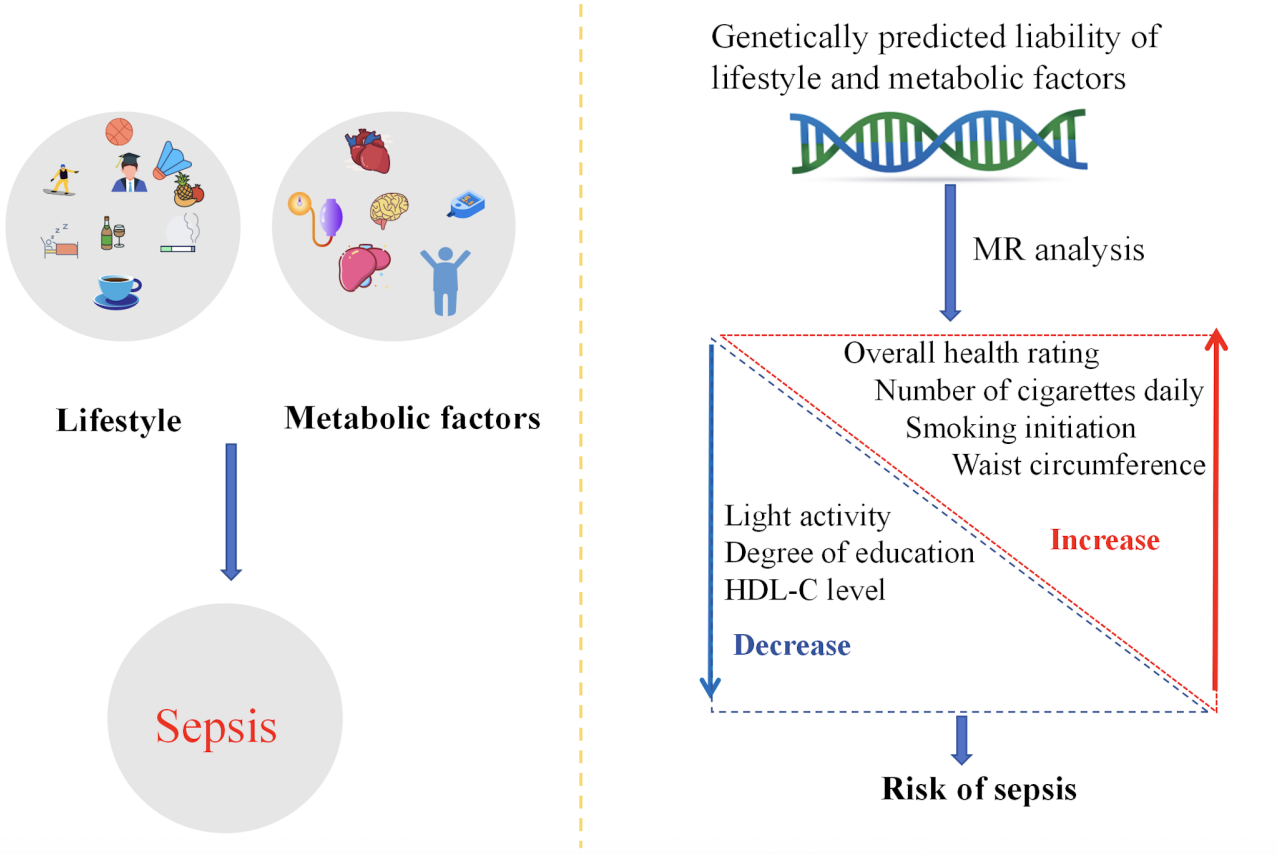
We conducted a sensitivity analysis using the MR-Egger intercept tests and the leave-one-out method. MR-Egger intercept tests showed no evidence of pleiotropy (Table 2). Using leave-one-out analysis, it is evident that nearly all lines align on one side of the y-axis, with minimal deviation beyond the axis. This consistency indicated the robustness of the results. The stable results shown in the MR visualization further corroborate the reliability of our study findings.

Discussion

In this bidirectional MR study, we investigated the association of lifestyle, educational and psychological, and metabolic factors with the risk of sepsis. Genetic variants such as smoking initiation, smoking frequency, and waist circumference were associated with an increased risk of sepsis, whereas light physical activity, higher education, and high levels of HDL-c were causally associated with a lower risk of sepsis (Figure 3).

XSL•FO
RenderX

Figure 3. Association of modifiable lifestyle and metabolic factors with the risk of developing sepsis. HDL-C: high-density lipoprotein cholesterol; MR: Mendelian randomization.



Genetic variants such as smoking initiation, smoking frequency, and waist circumference were associated with an increased risk of sepsis, whereas light physical activity, higher education, and high levels of HDL-c were causally associated with a lower risk of sepsis.

Owing to the modifiable nature of lifestyle, the impact of lifestyle on diseases with high mortality, such as sepsis, has received increasing attention. Numerous observational studies have described an association between smoking, alcohol consumption, and the risk of various infectious diseases [22,30-35]. However, establishing a causal link between lifestyle and sepsis remains challenging. This study aimed to explore the potential causal relationships between lifestyle factors and sepsis. The results showed that smoking initiation and a high frequency of smoking were associated with an increased risk of sepsis. Previous immunologic experiments reported that smoking has been associated with impaired phagocytic function and cytokine expression of T lymphocytes and macrophages in the respiratory tract or peripheral blood [36-38]. These studies collectively underscore the detrimental impact of smoking on the immune system. These findings provide a better understanding of the role of smoking in patients with sepsis. Therefore, healthy lifestyle modification is a preventive measure against sepsis; individuals should be advised to reduce smoking to reduce the risk of developing sepsis.

Physical exercise is a planned, purposeful action to maintain a healthy lifestyle and improve physical fitness (World Health Organization, 2010) [39]. Furthermore, the evidence suggests that physical exercise affects multiple organ systems under various conditions [40]. It has been found that inadequate exercise is associated with a doubled risk of sepsis-related mortality [40]. Given that the incidence of sepsis is increasing, these results suggest that exercising could be immediately effective in reducing disease incidence. Currently, no MR analyses of the association between physical activity and sepsis are available. Our analysis confirmed that light physical activity is associated with a lower risk of sepsis.

The socioeconomic status affects health through environmental exposure, health behaviors, and lifestyle and has been shown to be positively related to health [10,41]. Education stands as the most potent indicator of socioeconomic status, exerting influence over lifestyle choices and access to health resources [42]. Observational studies suggest an inverse association between educational attainment and sepsis risk. Wang et al [43] reported significantly higher sepsis risk among individuals with lower education (hazard ratio 1.88, 95% CI 1.54 - 2.29), though without controlling for comorbidities and other confounders. Another cohort study adjusting for socioeconomic factors still demonstrated a higher risk of sepsis-related intensive care unit admission in moderately educated versus highly educated individuals [44]. It should be noted that these observational findings may be influenced by residual confounding and do not establish causality. This study found a negative correlation between genetically predicted educational levels and sepsis risk. Higher education may reduce sepsis risk through multiple pathways, including enhanced health literacy (eg, earlier recognition of infections), greater access to health care resources, and adoption of healthier behaviors (eg, smoking

XSL•FO
RenderX

cessation and improved diet). These mechanisms could collectively mitigate exposure to risk factors and improve timely medical intervention.

In recent years, an increasing number of studies have focused on the impact of metabolic factors such as blood glucose, blood lipids, and obesity on various infectious diseases [45-49]. Previous studies have reported that high HDL-c levels are significantly associated with reduced sepsis risk and other sepsis-related outcomes [50,51]. Extremely low serum HDL-c levels (≤20 mg/dL) are associated with an increased risk of death, sepsis, and malignancy [52]. However, observational clinical and epidemiologic studies have potential problems such as confounding, reverse causation, and unaccounted comorbidities. An MR study found that lower HDL-c levels were significantly associated with an increased risk of sepsis and related outcomes in infected patients [53]. This is consistent with our findings. Our study found a genetic relationship between high HDL-c levels and low risk of sepsis. HDL-c might exert protective effects against sepsis via its anti-inflammatory and antioxidant properties, as well as its role in modulating endothelial function and neutralizing bacterial endotoxins [54-56].

In observational studies on the general population, a higher BMI has been associated with an increased incidence of mortality from bloodstream infections and sepsis [57-59]. This finding is confirmed in this study. We also found that BMI, arm fat mass, leg fat mass, whole body fat mass, waist circumference, and trunk fat mass were significantly associated with sepsis. However, using the multivariate factor analysis, only increased waist circumference was an independent risk factor for sepsis, which might be because other factors included the causes of increased waist circumference, suggesting that these factors might influence sepsis risk as a consequence of shared risk factor profiles.

This study had several limitations. First, the population limited to the European ancestry hampered the generalization of the findings to individuals of other ancestries, as genetic and environmental factors influencing sepsis risk may vary across ethnic groups (eg, differences in inflammatory responses or health care access). Further studies are required to assess the modifiable risk of sepsis in other races. Second, as with all MR studies, pleiotropy in an MR setting is challenging. Additionally, gene-environment interactions (eg, how lifestyle modifies genetic risk) and potential measurement errors in exposure variables were not fully addressed, which might have introduced bias. We performed various sensitivity analyses that made different assumptions regarding the underlying nature of pleiotropy. Most tests showed stable results. Future studies should incorporate multi-ancestry cohorts and explore gene-environment interplay to strengthen causal inference.

In conclusion, our combined MR analysis supports the causal roles of smoking, educational level, HDL-c level, physical condition, obesity, and physical activity in patients with sepsis. The findings of this study highlight actionable targets for sepsis prevention. For instance, smoking cessation programs, public health initiatives to promote physical activity, and interventions to improve HDL-c levels (eg, dietary modifications or pharmacotherapy) could be integrated into personalized preventive strategies for managing high-risk populations. Furthermore, educational campaigns targeting health literacy and early symptom recognition might reduce delays in seeking care, particularly among socioeconomically disadvantaged groups. Future research should prioritize replication in diverse populations, mechanistic studies to elucidate pathways (eg, HDL-c's immunomodulatory effects), and translational interventions targeting high-risk subgroups.

Acknowledgments

The authors would like to acknowledge the participants and investigators of the Integrative Epidemiology Unit OpenGWAS project used in this paper.

Data Availability

Publicly available datasets were analyzed in this Mendelian randomization study. Genome-wide association study summary-level datasets were downloaded from the Integrative Epidemiology Unit OpenGWAS Project [29]. Full details are provided in Table 1.

Authors' Contributions

Data curation: HL, JL
Formal analysis: HL, JL
Methodology: YC
Project administration: KX
Software: HL, JL
Supervision: XW, KX
Validation: GS, FW, QY
Visualization: HL, JL, WF
Writing–original draft: HL, JL, YC, KX, XW, GS, FW, QY, WF
Writing–review and editing: HL, JL, YC, KX, XW, GS, FW, QY, WF

XSL•FO
RenderX

Conflicts of Interest

None declared.

Checklist 1
STROBE-MR checklist.
[DOCX File, 31 KB - i-jmr_v14i1e72244_app1.docx]

References

1. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 2016 Feb 23;315(8):801-810. [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
2. 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]
3. Xie J, Wang H, Kang Y, et al. The epidemiology of sepsis in Chinese ICUs: a national cross-sectional survey. Crit Care Med 2020 Mar;48(3):e209-e218. [doi: 10.1097/CCM.0000000000004155] [Medline: 31804299]
4. Global report on the epidemiology and burden of sepsis: current evidence, identifying gaps and future directions. : World Health Organization; 2020 URL: https://www.who.int/publications/i/item/9789240010789 [accessed 2025-10-23]
5. Vincent JL, van der Poll T, Marshall JC. The end of “one size fits all” sepsis therapies: toward an individualized approach. Biomedicines 2022 Sep 12;10(9):2260. [doi: 10.3390/biomedicines10092260] [Medline: 36140361]
6. 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]
7. Vincent JL. Emerging paradigms in sepsis. EBioMedicine 2022 Dec;86:104398. [doi: 10.1016/j.ebiom.2022.104398] [Medline: 36470835]
8. Rubio I, Osuchowski MF, Shankar-Hari M, et al. Current gaps in sepsis immunology: new opportunities for translational research. Lancet Infect Dis 2019 Dec;19(12):e422-e436. [doi: 10.1016/S1473-3099(19)30567-5]
9. Wiewel MA, Harmon MB, van Vught LA, et al. Risk factors, host response and outcome of hypothermic sepsis. Crit Care 2016 Oct 14;20(1):328. [doi: 10.1186/s13054-016-1510-3] [Medline: 27737683]
10. Stensrud VH, Gustad LT, Damås JK, Solligård E, Krokstad S, Nilsen TIL. Direct and indirect effects of socioeconomic status on sepsis risk and mortality: a mediation analysis of the HUNT study. J Epidemiol Community Health 2023 Mar;77(3):168-174. [doi: 10.1136/jech-2022-219825] [Medline: 36707239]
11. Rose N, Matthäus-Krämer C, Schwarzkopf D, et al. Association between sepsis incidence and regional socioeconomic deprivation and health care capacity in Germany - an ecological study. BMC Public Health 2021 Sep 7;21(1):1636. [doi: 10.1186/s12889-021-11629-4] [Medline: 34493250]
12. Trevelin SC, Carlos D, Beretta M, da Silva JS, Cunha FQ. Diabetes mellitus and sepsis: a challenging association. Shock 2017 Mar;47(3):276-287. [doi: 10.1097/SHK.0000000000000778] [Medline: 27787406]
13. Ahlberg CD, Wallam S, Tirba LA, Itumba SN, Gorman L, Galiatsatos P. Linking Sepsis with chronic arterial hypertension, diabetes mellitus, and socioeconomic factors in the United States: a scoping review. J Crit Care 2023 Oct;77:154324. [doi: 10.1016/j.jcrc.2023.154324] [Medline: 37159971]
14. Stösser S, Kleusch L, Schenk A, Schmid M, Petzold GC. Derivation and validation of a screening tool for stroke-associated sepsis. Neurol Res Pract 2023 Jul 13;5(1):32. [doi: 10.1186/s42466-023-00258-4] [Medline: 37438794]
15. Hou J, Zhang J, Cui P, et al. TREM2 sustains macrophage-hepatocyte metabolic coordination in nonalcoholic fatty liver disease and sepsis. J Clin Invest 2021 Feb 15;131(4):e135197. [doi: 10.1172/JCI135197] [Medline: 33586673]
16. You J, Bi X, Zhang K, et al. Causal associations between gut microbiota and sepsis: a two-sample Mendelian randomization study. Eur J Clin Invest 2023 Nov;53(11):e14064. [doi: 10.1111/eci.14064] [Medline: 37464539]
17. Wang GS, You KM, Jo YH, et al. Association of health insurance status with outcomes of sepsis in adult patients: a retrospective cohort study. Int J Environ Res Public Health 2021 May 27;18(11):5777. [doi: 10.3390/ijerph18115777] [Medline: 34072210]
18. Frydrych LM, Bian G, O’Lone DE, Ward PA, Delano MJ. Obesity and type 2 diabetes mellitus drive immune dysfunction, infection development, and sepsis mortality. J Leukoc Biol 2018 Sep;104(3):525-534. [doi: 10.1002/JLB.5VMR0118-021RR] [Medline: 30066958]
19. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ 2018;362:k601. [doi: 10.1136/bmj.k601]
20. Lei P, Xu W, Wang C, Lin G, Yu S, Guo Y. Mendelian randomization analysis reveals causal associations of polyunsaturated fatty acids with sepsis and mortality risk. Infect Dis Ther 2023 Jul;12(7):1797-1808. [doi: 10.1007/s40121-023-00831-z] [Medline: 37316614]
21. Wang J, Hu Y, Zeng J, et al. Exploring the causality between body mass index and sepsis: a two-sample Mendelian randomization study. Int J Public Health 2023;68:1605548. [doi: 10.3389/ijph.2023.1605548] [Medline: 37205044]

XSL•FO
RenderX

22. Zhu H, Zhan X, Wang C, et al. Causal associations between tobacco, alcohol use and risk of infectious diseases: a Mendelian randomization study. Infect Dis Ther 2023 Mar;12(3):965-977. [doi: 10.1007/s40121-023-00775-4] [Medline: 36862322]
23. Hu J, Gan Q, Zhou D, et al. Evaluating the risk of sepsis attributing to obesity: a two-sample Mendelian randomization study. Postgrad Med J 2023 Nov 20;99(1178):1266-1271. [doi: 10.1093/postmj/qgad072] [Medline: 37681245]
24. Burgess S, Davey Smith G, Davies NM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. Wellcome Open Res 2019;4:186. [doi: 10.12688/wellcomeopenres.15555.3] [Medline: 32760811]
25. Thorkildsen MS, Gustad LT, Mohus RM, et al. Association of genetically predicted insomnia with risk of sepsis: a Mendelian randomization study. JAMA Psychiatry 2023 Oct 1;80(10):1061-1065. [doi: 10.1001/jamapsychiatry.2023.2717] [Medline: 37556136]
26. Ni F, Liu X, Wang S. Impact of negative emotions and insomnia on sepsis: a mediation Mendelian randomization study. Comput Biol Med 2024 Sep;180:108858. [doi: 10.1016/j.compbiomed.2024.108858] [Medline: 39067155]
27. Shang W, Qian H, Zhang S, et al. Human blood metabolites and risk of sepsis: a Mendelian randomization investigation. Eur J Clin Invest 2024 Apr;54(4):e14145. [doi: 10.1111/eci.14145] [Medline: 38041600]
28. Zhu ZG, Ma JW, Ji DD, Li QQ, Diao XY, Bao J. Mendelian randomization analysis identifies causal associations between serum lipidomic profile, amino acid biomarkers and sepsis. Heliyon 2024 Jun 30;10(12):e32779. [doi: 10.1016/j.heliyon.2024.e32779] [Medline: 38975226]
29. OpenGWAS. URL: https://opengwas.io/ [accessed 2025-10-28]
30. O'Brien JM Jr, Lu B, Ali NA, et al. Alcohol dependence is independently associated with sepsis, septic shock, and hospital mortality among adult intensive care unit patients. Crit Care Med 2007 Feb;35(2):345-350. [doi: 10.1097/01.CCM.0000254340.91644.B2] [Medline: 17205003]
31. Bello S, Menéndez R, Antoni T, et al. Tobacco smoking increases the risk for death from pneumococcal pneumonia. Chest 2014 Oct;146(4):1029-1037. [doi: 10.1378/chest.13-2853] [Medline: 24811098]
32. Gustot T, Fernandez J, Szabo G, et al. Sepsis in alcohol-related liver disease. J Hepatol 2017 Nov;67(5):1031-1050. [doi: 10.1016/j.jhep.2017.06.013] [Medline: 28647569]
33. Bukong TN, Cho Y, Iracheta-Vellve A, et al. Abnormal neutrophil traps and impaired efferocytosis contribute to liver injury and sepsis severity after binge alcohol use. J Hepatol 2018 Nov;69(5):1145-1154. [doi: 10.1016/j.jhep.2018.07.005] [Medline: 30030149]
34. Moazed F, Hendrickson C, Jauregui A, et al. Cigarette smoke exposure and acute respiratory distress syndrome in sepsis: epidemiology, clinical features, and biologic markers. Am J Respir Crit Care Med 2022 Apr 15;205(8):927-935. [doi: 10.1164/rccm.202105-1098OC] [Medline: 35050845]
35. Sabaratnam R, Wojtaszewski JFP, Højlund K. Factors mediating exercise-induced organ crosstalk. Acta Physiol (Oxf) 2022 Feb;234(2):e13766. [doi: 10.1111/apha.13766] [Medline: 34981891]
36. Burton R, Fryers PT, Sharpe C, et al. The independent and joint risks of alcohol consumption, smoking, and excess weight on morbidity and mortality: a systematic review and meta-analysis exploring synergistic associations. Public Health 2024 Jan;226:39-52. [doi: 10.1016/j.puhe.2023.10.035] [Medline: 38000113]
37. Lee EH, Lee KH, Lee KN, Park Y, Do Han K, Han SH. The relation between cigarette smoking and development of sepsis: a 10-year follow-up study of four million adults from the national health screening program. J Epidemiol Glob Health 2024 Jun;14(2):444-452. [doi: 10.1007/s44197-024-00197-6] [Medline: 38372892]
38. Elisia I, Lam V, Cho B, et al. The effect of smoking on chronic inflammation, immune function and blood cell composition. Sci Rep 2020 Nov 10;10(1):19480. [doi: 10.1038/s41598-020-76556-7] [Medline: 33173057]
39. Ye KX, Sun L, Wang L, et al. The role of lifestyle factors in cognitive health and dementia in oldest-old: a systematic review. Neurosci Biobehav Rev 2023 Sep;152:105286. [doi: 10.1016/j.neubiorev.2023.105286] [Medline: 37321363]
40. Williams PT. Inadequate exercise as a risk factor for sepsis mortality. PLoS One 2013;8(12):e79344. [doi: 10.1371/journal.pone.0079344] [Medline: 24324580]
41. Minejima E, Wong-Beringer A. Impact of socioeconomic status and race on sepsis epidemiology and outcomes. J Appl Lab Med 2021 Jan 12;6(1):194-209. [doi: 10.1093/jalm/jfaa151] [Medline: 33241269]
42. Rosengren A, Smyth A, Rangarajan S, et al. Socioeconomic status and risk of cardiovascular disease in 20 low-income, middle-income, and high-income countries: the Prospective Urban Rural Epidemiologic (PURE) study. Lancet Glob Health 2019 Jun;7(6):e748-e760. [doi: 10.1016/S2214-109X(19)30045-2] [Medline: 31028013]
43. Wang HE, Shapiro NI, Griffin R, Safford MM, Judd S, Howard G. Chronic medical conditions and risk of sepsis. PLoS One 2012;7(10):e48307. [doi: 10.1371/journal.pone.0048307] [Medline: 23118977]
44. Storm L, Schnegelsberg A, Mackenhauer J, Andersen LW, Jessen MK, Kirkegaard H. Socioeconomic status and risk of intensive care unit admission with sepsis. Acta Anaesthesiol Scand 2018 Aug;62(7):983-992. [doi: 10.1111/aas.13114] [Medline: 29569230]
45. Madsen CM, Varbo A, Tybjærg-Hansen A, Frikke-Schmidt R, Nordestgaard BG. U-shaped relationship of HDL and risk of infectious disease: two prospective population-based cohort studies. Eur Heart J 2018 Apr 7;39(14):1181-1190. [doi: 10.1093/eurheartj/ehx665] [Medline: 29228167]

XSL•FO
RenderX

46. Meng J, Li X, Xiao Y, et al. Intensive or liberal glucose control in intensive care units for septic patients? A meta-analysis of randomized controlled trials. Diabetes Metab Syndr 2024 May;18(5):103045. [doi: 10.1016/j.dsx.2024.103045] [Medline: 38796958]
47. Liu P, Meng J, Tang H, et al. Association between bariatric surgery and outcomes of total joint arthroplasty: a meta-analysis. Int J Surg 2025;111(1):1541-1546. [doi: 10.1097/JS9.0000000000002002]
48. Meng J, Li X, Liu W, et al. The role of vitamin D in the prevention and treatment of SARS-CoV-2 infection: a meta-analysis of randomized controlled trials. Clin Nutr 2023 Nov;42(11):2198-2206. [doi: 10.1016/j.clnu.2023.09.008] [Medline: 37802017]
49. Meng J, Li X, Xiong Y, Wu Y, Liu P, Gao S. The role of vitamin D in the prevention and treatment of tuberculosis: a meta-analysis of randomized controlled trials. Infection 2025 Jun;53(3):1129-1140. [doi: 10.1007/s15010-024-02446-z]
50. Scalsky RJ, Chen YJ, Desai K, O'Connell JR, Perry JA, Hong CC. Baseline cardiometabolic profiles and SARS-CoV-2 infection in the UK Biobank. PLoS ONE 2021;16(4):e0248602. [doi: 10.1371/journal.pone.0248602]
51. Shor R, Wainstein J, Oz D, et al. Low HDL levels and the risk of death, sepsis and malignancy. Clin Res Cardiol 2008 Apr;97(4):227-233. [doi: 10.1007/s00392-007-0611-z]
52. Liu G, Jiang L, Kerchberger VE, et al. The relationship between high density lipoprotein cholesterol and sepsis: a clinical and genetic approach. Clinical Translational Sci 2023 Mar;16(3):489-501. [doi: 10.1111/cts.13462]
53. Twig G, Geva N, Levine H, et al. Body mass index and infectious disease mortality in midlife in a cohort of 2.3 million adolescents. Int J Obes (Lond) 2018 Apr;42(4):801-807. [doi: 10.1038/ijo.2017.263] [Medline: 29081504]
54. Hamilton F, Pedersen KM, Ghazal P, Nordestgaard BG, Smith GD. Low levels of small HDL particles predict but do not influence risk of sepsis. Crit Care 2023 Oct 9;27(1):389. [doi: 10.1186/s13054-023-04589-1] [Medline: 37814277]
55. Taylor R, Zhang C, George D, et al. Low circulatory levels of total cholesterol, HDL-C and LDL-C are associated with death of patients with sepsis and critical illness: systematic review, meta-analysis, and perspective of observational studies. EBioMedicine 2024 Feb;100:104981. [doi: 10.1016/j.ebiom.2024.104981] [Medline: 38290288]
56. Zou G, Zhu Q, Ren B, et al. HDL-associated lipoproteins: potential prognostic biomarkers for gram-negative sepsis. J Inflamm Res 2022;15:1117-1131. [doi: 10.2147/JIR.S350737] [Medline: 35210815]
57. Weng L, Fan J, Yu C, et al. Body-mass index and long-term risk of sepsis-related mortality: a population-based cohort study of 0.5 million Chinese adults. Crit Care 2020 Aug 31;24(1):534. [doi: 10.1186/s13054-020-03229-2] [Medline: 32867859]
58. Paulsen J, Askim Å, Mohus RM, et al. Associations of obesity and lifestyle with the risk and mortality of bloodstream infection in a general population: a 15-year follow-up of 64 027 individuals in the HUNT Study. Int J Epidemiol 2017 Oct 1;46(5):1573-1581. [doi: 10.1093/ije/dyx091] [Medline: 28637260]
59. Winter-Jensen M, Afzal S, Jess T, Nordestgaard BG, Allin KH. Body mass index and risk of infections: a Mendelian randomization study of 101,447 individuals. Eur J Epidemiol 2020 Apr;35(4):347-354. [doi: 10.1007/s10654-020-00630-7] [Medline: 32307655]

Abbreviations

GWAS: genome-wide association study
HDL-C: high-density lipoprotein cholesterol
IVW: inverse variance weighted
MR: Mendelian randomization
OR: odds ratio
SNP: single-nucleotide polymorphism
STROBE-MR: Strengthening the Reporting of Observational Studies in Epidemiology using Mendelian Randomization

Edited by N Cahill; submitted 06.02.25; peer-reviewed by B He, O Abasilim; revised version received 01.10.25; accepted 01.10.25; published 03.11.25.

Please cite as:
Lv H, Liu J, Cao Y, Fan W, Shen G, Wang F, Ye Q, Wu X, Xu K
Association of Modifiable Lifestyle and Metabolic Factors With the Risk of Developing Sepsis: 2-Sample Mendelian Randomized Study
Interact J Med Res 2025;14:e72244
URL: https://www.i-jmr.org/2025/1/e72244
doi:10.2196/72244

XSL•FO
RenderX

© Haifeng Lv, Jing Liu, Yelin Cao, Weina Fan, Guojie Shen, Feifei Wang, Qingqing Ye, Xiaoliang Wu, Kaijin Xu. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 3.11.2025. 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

Correlation Between Objective Habit Metrics and Objective Medication Adherence: Retrospective Study of 15,818 Participants From Clinical Studies

Antoine Pironet[1], PhD; L Alison Phillips[2], PhD; Bernard Vrijens[1], Prof Dr

[1]AARDEX Group, Seraing, Belgium

[2]Iowa State University, Ames, IA, United States

Corresponding Author:
Antoine Pironet, PhD
AARDEX Group
rue Bois St Jean 15/1
Seraing, 4102
Belgium
Phone: 32 43748640
Email: antoine.pironet@aardexgroup.com

Abstract

Background: Medication adherence, or how patients take their medication as prescribed, is suboptimal worldwide. Improving medication-taking habit might be an effective way to improve medication adherence. However, habit is difficult to quantify, and conventional habit metrics are self-reported, with recognized limitations. Recently, several objective habit metrics have been proposed, based on objective medication-taking data.

Objective: We aim to explore the correlation between objective habit metrics and objective medication adherence on a large dataset.

Methods: The Medication Event Monitoring System Adherence Knowledge Center, a database of anonymized electronic medication intake data from ambulant participants enrolled in past clinical studies, was used as the data source. Electronic medication intake data from participants following a once-daily regimen and monitored for 14 days or more were used. Further, two objective habit metrics were computed from each participant's medication intake history: (1) SD of the hour of intake, representing daily variability in the timing of medication intakes, and (2) weekly cross-correlation, representing weekly consistency in the timing of medication intakes. The implementation component of medication adherence was quantified using (1) the proportion of doses taken and (2) the proportion of correct days.

Results: A total of 15,818 participants met the criteria. These participants took part in 108 clinical studies mainly focused on treatments for hypertension (n=4737, 30%) and osteoporosis (n=3353, 21%). The SD of the hour of intake was significantly negatively correlated with the 2 objective adherence metrics: proportion of correct days (Spearman correlation coefficient, ρ_S=–0.62, P<.001) and proportion of doses taken (ρ_S=–0.09, P<.001). The weekly cross-correlation was significantly positively correlated with the 2 objective adherence metrics: proportion of correct days (ρ_S=0.55, P<.001) and proportion of doses taken (ρ_S=0.32, P<.001). A lower daily or weekly variability in the timing of medication intakes is thus associated with better medication adherence. However, no variability is not the norm, as only 3.6% of participants have 95% of their intakes in a 1-hour window. Among the numerous factors influencing medication adherence, habit strength is an important one as it explains over 30% of the variance in medication adherence.

Conclusions: Objective habit metrics are correlated to objective medication adherence. Such objective habit metrics can be used to monitor patients and identify those who may benefit from habit-building support.

(Interact J Med Res 2025;14:e63987) doi:10.2196/63987

KEYWORDS

medication adherence; compliance; habit; history; correlation; association; intake; electronic database; retrospective; medication; drug; adherence

XSL•FO
RenderX

Introduction

Medication adherence is "the process by which patients take their medication as prescribed" [1]. It comprises 3 components: initiation of the treatment, correct implementation of the prescribed regimen, and persistence to the treatment [1]. Poor medication adherence is a global public health issue [2,3] and has important negative consequences on the personal level [3-6], but also at the societal level [3-5]. Interventions aiming at improving medication adherence are abundant in the literature, but few have been shown to be effective across a population [5,6]. Most medication adherence interventions have focused on structural factors outside of the individual (simplification of the regimen, refill reminders, etc) or on behavior change interventions that target reflective or deliberative factors, such as patient education [7]. As at least half of our daily behaviors are nonreflective, but rather habitual [8], interventions that target these habitual processes may be more successful than education or persuasion-based interventions.

Habits are defined as automatic behaviors responding to recurring environmental cues [9]. As stronger medication-taking habit has been shown to be associated with better medication adherence [4,10-13], in particular its implementation component, some successful interventions have focused on improving habit [5,14]. However, medication-taking habit, and habit in general, is difficult to quantify [8]. Conventional habit metrics are self-reported [12], with an important example being the Self-Reported Habit Index [15], a 12-item questionnaire. Another example is the Self-Reported Behavioral Automaticity Index [16], a 4-item subset of the Self-Reported Habit Index. These and other self-reported indices endure common limitations associated with self-reported metrics, such as social desirability bias and poor patient recall [13].

Medication-taking habit can also be assessed using objective medication-taking data [17]. Such data is collected using smart medication packages, which can take several forms: an electronic cap fitted on a medication bottle [17], an inhaler with a chip embedded [13], a blister that detects when a pill is expressed out of a cavity, etc. The common feature of smart medication packages is that they passively timestamp each time a patient accesses their medication, thus providing objective data on when a patient takes their medication. This detailed information can be used to derive habit metrics quantifying the consistency of medication intake behavior over time [11,17]—the validity of which rests on the fact that habits are context-stable responses to conditioned cues. Such habit metrics are objective and do not endure the limitations associated with self-reported habit.

Day-to-day consistency of the timing of medication intake is frequently used as an objective habit metric. It has been operationalized as the variance of the hour of intake [17], its SD [13,18], or the proportion of medication intakes occurring in a fixed-size window, for instance, 2 [4,19,20], 3 [21,22], or 4 hours [23].

These day-to-day consistency measures will penalize someone for having different routines on different days of the week (for instance, systematically taking their medication at 6 PM on weekdays and 10 PM on weekends). However, this feature might not be desirable if the cue for medication-taking (eg, breakfast) is the same throughout the week, reflecting a good medication-taking habit. To overcome this limitation, Phillips et al [11] recently introduced another metric, called the weekly cross-correlation, based on the weekly consistency of intake timing. In this metric, the medication intake timing of each day of a given week is compared to the corresponding day of the next week.

Recently also, Hoo et al [13] introduced a "pragmatic habit index" empirically defined as the product of 2 variables: stability, measured as the SD of the hour of intake and frequency, measured as the proportion of prescribed doses taken. Introducing behavioral frequency in a habit index has been criticized [16], as it incorporates the dependent variable of interest (behavior frequency) in the predictors.

In a recent study on 79 patients with type 2 diabetes, objective habit metrics were found to correlate with objective medication adherence [11]. The goal of this paper is to reinvestigate this correlation on a 200-fold larger dataset covering multiple pathologies, introduced in the next section.

The main hypothesis is that a more consistent medication-taking habit is correlated to better medication adherence.

Methods

Data

This study was a retrospective cohort study. AARDEX Group's database, the Medication Event Monitoring System (MEMS; AARDEX Group) Adherence Knowledge Center was used as the data source. This database contains anonymized electronic medication adherence data from ambulant participants enrolled in clinical studies that ran between 1989 and 2016. These participants' medication adherence was electronically monitored using the MEMS. The following selection criteria were used: (1) once-daily medication with (2) a follow-up longer than 14 days. This second criterion was required because 14 days is the minimum duration needed to compute the weekly cross-correlation. These selection criteria matched data from 15,818 participants totaling 3,053,779 medication intakes.

Ethical Considerations

Human Participants' Ethics Review Approval or Exemptions

Approval from an institutional review board was not required for the present analysis, as it consists of secondary research for which the identity of the participants is unknown, per Title 45 of the Code of Federal Regulations Part 46, Subpart A, Section 46.104, Paragraph (d)(4)(ii) [24]. The original data collection for all studies included in this analysis was approved by institutional review boards.

Informed Consent

Data were obtained from participants enrolled in clinical studies. As such, the participants provided informed consent for the procedures of the original studies. These procedures required participants to store their medication in electronic medication

XSL•FO
RenderX

packages. Participants were informed that these electronic medication packages recorded their medication intakes. The original study procedures also included using the generated data to monitor the participants' dosing history and analyze their medication intake behavior. As a consequence, informed consent was not sought for this secondary analysis, because it corresponds to the use of the data that was originally presented to participants.

Privacy and Confidentiality

The data in AARDEX Group's database, the Adherence Knowledge Center, is anonymized.

Compensation Details:

Participants were compensated for their participation in the original clinical studies. Compensation modalities varied between studies and countries, and the authors do not possess information about the compensation process, which was managed by the sponsors of the original studies.

Objective Habit Metrics

For each participant, 2 objective habit metrics were computed. The first objective habit metric is the SD of the day-to-day hour of intake. The hours and minutes are extracted for each medication intake timestamp, irrespective of the date of intake, and the SD of the resulting list of hours and minutes is computed. The lower the SD, the more consistent a person is in the timing of their medication intakes. In the extreme case, when SD equals 0 h, all intakes occur at exactly the same time of the day.

The second objective habit metric is the "weekly cross-correlation" introduced by Phillips et al [11]. This metric compares each day of a given week with the corresponding day of the next week and quantifies whether intakes occurred around the same time on these 2 days. To do so, medication intakes are represented in a 2D matrix, denoted A, of dimensions $24 \times N_{days}$, with N_{days} being the number of follow-up days. The matrix is initially filled with zeros. Each medication intake occurring on day l at hour k is translated as a value of 1 for element $A[k,l]$. Then, each element $A[k,l]$ of the matrix is multiplied by its matching element from the previous week, $A[k,l-7]$. If intakes occur at the same time from week to week, the product will be close to 1. Otherwise, it will be close to zero. Finally, the weekly cross-correlation is equal to the sum of all element-wise products divided by the Euclidean norm of A. More details and code explaining how to compute these metrics are presented in Multimedia Appendix 1. As the pragmatic habit index incorporates adherence, it was not included in the present analysis.

Implementation Adherence Metrics

Medication adherence, more precisely its implementation component [1], was quantified as the proportion of correct days, that is, the proportion of days with exactly 1 intake. A second implementation metric was used: the proportion of doses taken. The proportion of doses taken was computed as the ratio between the total number of doses taken and the prescribed number of doses. In this work, since participants were on a once-daily regimen, the prescribed number of doses was equal to the number of follow-up days. If a participant discontinued treatment too early, the analysis was limited to the period during which the participant was on treatment.

The hypotheses were that (1) SD of the day-to-day hour of intake would be negatively correlated to objective adherence (the more variable the hour of intake, the worse the adherence); (2) the weekly cross-correlation would be positively correlated to objective adherence (the more consistent the pattern of intakes, the better the adherence); and (3) the weekly cross-correlation would be more strongly correlated to objective adherence than the SD of the hour of intake.

All dosing history data were compiled while patients were engaged with the monitored medication in the trial. Therefore, noninitiation and nonpersistence are not part of the adherence evaluation in this research.

Statistical Analyses

Continuous variables were reported using medians and first and third quartiles. Categorical variables were reported using counts and proportions. Pairwise correlations between continuous variables were assessed using the Spearman rank correlation coefficient. Correlation coefficients were converted using Fisher Z-transformation to obtain CIs [25,26]. Correlation coefficients were compared by performing a z-test on the difference between their Fisher Z-transformations [26,27]. The proportion of variance in adherence explained by habit was quantified using the Pearson correlation coefficient, squared [26]. Analyses were performed using Python 3 (Python Software Foundation) [28].

Results

The 15,818 participants whose data was extracted from the MEMS Adherence Knowledge Center took part in 108 clinical studies. These studies enrolled a median of 59 (IQR 22-169) participants. Table 1 presents the characteristics of the population.

Figure 1 presents the distributions and pairwise plots for the 4 variables of interest: proportion of correct days and proportion of doses taken as 2 objective measures of adherence, and SD of the hour of intake and weekly cross-correlation as 2 objective habit metrics. According to the 2 top-left diagonal panels, the proportion of correct days ranges between 0 and 1 by definition, while the proportion of doses taken ranges between 0 and 1.5 indicating that some participants took more doses than prescribed. The scatter plots also show that the proportion of doses taken is correlated to the proportion of correct days. In addition, the first is always superior to the second, which is a consequence of their definition.

According to the 2 bottom-right diagonal panels of Figure 1, the SD of the hour of intake ranges between 0 and 12 hours, while the weekly cross-correlation ranges between 0 and 1. These two findings are direct consequences of the definitions of these variables. According to the scatter plots, these 2 variables are negatively correlated, indicating that a larger consistency in day-to-day timing (low SD hour of intake) is associated with a larger consistency in week-to-week timing (high weekly cross-correlation).

XSL•FO
RenderX

Table 2 presents the associated correlation coefficients. All 4 Spearman correlation coefficients were significantly different from zero (T scores for the significance of the coefficients, from left to right and top to bottom: –61.48, –10.70, 104.11, and 48.59 with a df of 15,816, all P<.001).

According to Figure 1 and Table 2, the correlation between the SD of the hour of intake and objective adherence (measured using the proportion of correct days or proportion of doses taken) has a negative sign, indicating that, as hypothesized, the more day-to-day consistency in timing the higher the adherence. Figure 1 and Table 2 also show that the weekly cross-correlation is positively correlated to objective adherence, meaning that the more consistent a person's medication intake pattern is from week to week, the higher their medication adherence, which is also in line with the hypotheses.

The comparison of the Spearman correlation coefficients presented in Table 2 shows that the SD of the hour of intake is more strongly correlated to the proportion of correct days than the weekly cross-correlation is (95% CI for the difference between absolute values of Spearman coefficients: 0.05 to 0.09, z score for this difference=21.46, P<.001), conversely to what was hypothesized. On the other hand, the weekly cross-correlation is more strongly correlated to the proportion of doses taken than the SD hour of intake is (CI for the difference between absolute values of Spearman coefficients: 0.25 to 0.21, z score for this difference=9.45, P<.001), as hypothesized.

Habit accounts for about 30% of the variance in medication adherence, measured using the proportion of correct days: 29.18% for SD of the hour of intake (95% CI 29.17% to 29.19%) and 31.25% for the weekly cross-correlation (95% CI 31.24% to 31.26%). When medication adherence is measured using the proportion of doses taken, the proportion of variance explained by habit is smaller: 4.91% for the SD of the hour of intake (95% CI 4.91% to 4.92%) and 15.74% for the weekly cross-correlation (95% CI 15.73% to 15.75%).

Table 1. Population characteristics.

Pathology	Values
Hypertension, n (%)	4737 (30)
Osteoporosis, n (%)	3353 (21)
Viral hepatitis, n (%)	2005 (13)
Hypercholesterolemia, n (%)	1246 (8)
Angina, n (%)	968 (6)
AIDS, n (%)	754 (5)
Depression, n (%)	561 (4)
Diabetes, n (%)	549 (3)
Reversible airway obstruction, n (%)	397 (3)
Heart failure, n (%)	334 (2)
Attention-deficit/hyperactivity disorder, n (%)	301 (2)
Colorectal polyp, n (%)	192 (1)
Others, n (%)	421 (3)
Follow-up duration (days), median (IQR)	168 (60-365)
Proportion of correct days (%), median (IQR)	93.3 (83.7-97.7)
Proportion of doses taken (%), median (IQR)	100 (93.7-101.8)
SD hour of intake (h), median (IQR)	1.7 (1-2.7)
Weekly cross-correlation, median (IQR)	0.54 (0.38-0.69)

XSL•FO
RenderX

Figure 1. Pairwise relationships between the four variables of interest: 2 objective adherence metrics: proportion of correct days (number 1) and proportion of doses taken (number 2) and 2 objective habit metrics (numbers 3 and 4). Off-diagonal panels contain scatter plots for each pair of variables; on-diagonal panels show a histogram of the distribution of single variables.

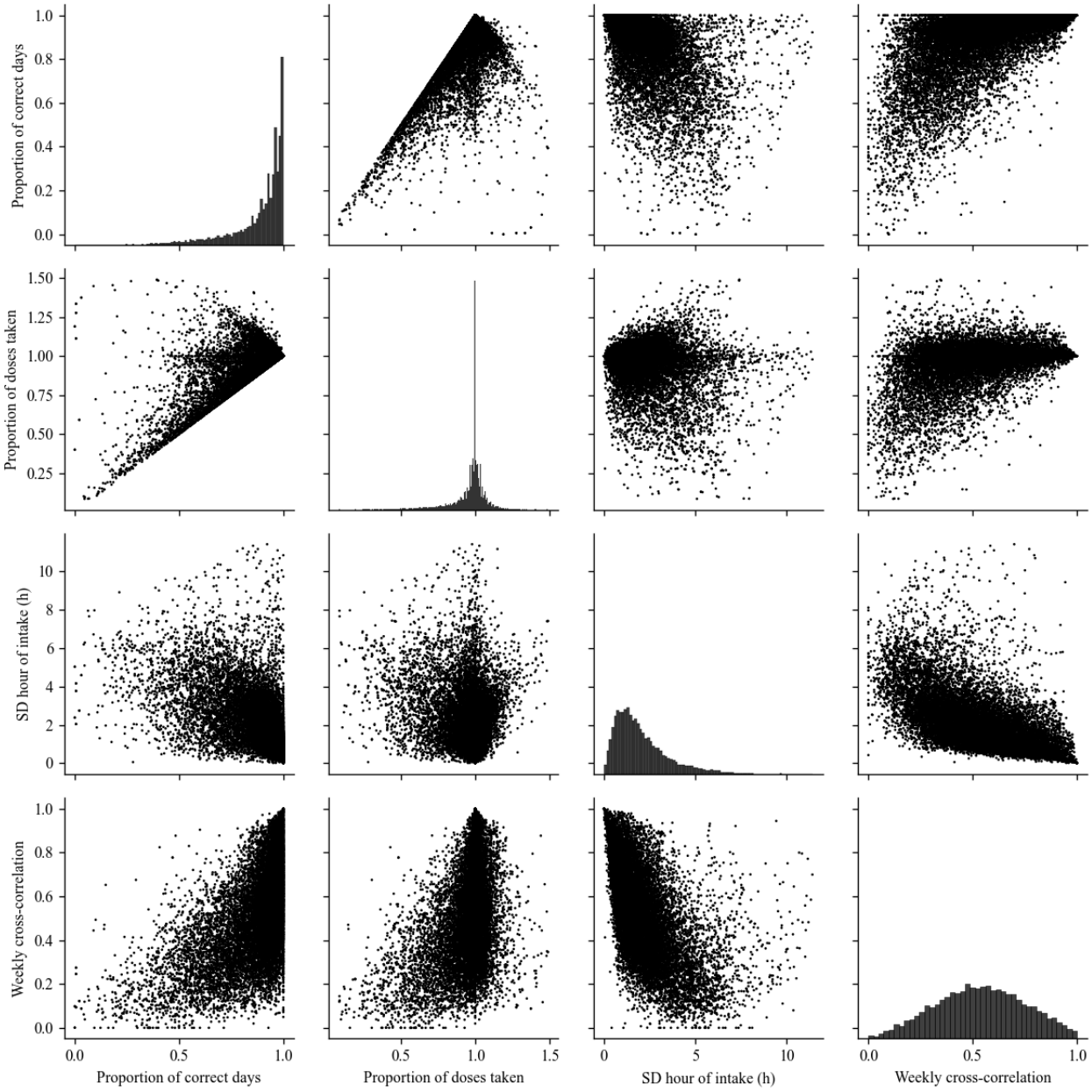


Table 2. Spearman correlation coefficients with 95% CI between 2 objective habit metrics (rows) and 2 objective adherence metrics (columns).

Objective habit	Objective adherence	
	Proportion of correct days	Proportion of doses taken
SD hour of intake	–0.62 (–0.63 to –0.61)	–0.09 (–0.1 to –0.07)
Weekly cross-correlation	0.55 (0.54 to 0.56)	0.32 (0.3 to 0.33)

Discussion

Principal Findings

This work showed that medication-taking habit is positively correlated to medication adherence in a very large, cross-pathology population of participants enrolled in clinical studies. Medication-taking habit was quantified using 2 measures of the consistency in the timing of medication intakes: day-to-day and week-to-week. The implementation component of medication adherence was quantified using 2 measures: the proportion of correct days and the proportion of doses taken. The conclusion was the same irrespective of the habit or adherence measure: the stronger the habit, the better the medication adherence.

XSL•FO
RenderX

The results showed no clear superiority of any of the 2 objective habit metrics per the strength of their correlation with adherence. The correlation of objective habit metrics with self-reported habit metrics, which are the most frequently used habit metrics, might be used to differentiate the 2 objective habit metrics. Previous studies exploring this correlation are discussed in the next section.

Comparison to Prior Work

The link between objective habit metrics and medication adherence has been investigated in very few other studies. The first one [11], performed by the authors on a much smaller dataset, reported very similar results: (1) a negative correlation between a measure of day-to-day variability and the proportion of correct days and (2) a positive correlation between the weekly cross-correlation and the proportion of correct days. The second one [13], previously discussed, used a habit metric incorporating adherence and reported an expected correlation between these 2 quantities.

Over 700 determinants of adherence have been reported [29]. Among all these determinants, objective habit strength seems to be an important one, as it explained over 30% of the variance in medication adherence in this work. This observation empirically justifies studying objective medication-taking habit.

A question related to the present work is whether people use time-based cues for taking their medications [10]. To answer this question, the proportion of participants having 95% of their intakes in a 1-hour window was computed. Only 3.6% of participants met this criterion. In a previous study [10], 21.7% of participants reported taking their medication at a specific time of the day. These numbers mean that a minority of patients rely on cues triggered by the clock and take medication at the same time every day. This finding might explain why reminders for medication adherence sent at a specific time of the day can be ineffective [30].

Few studies have studied how well objective habit metrics for medication adherence correlate with conventional self-reported habit metrics. SD or other metrics of day-to-day consistency were found to correlate with self-reported habit in 2 studies [17,18], but not in a third one [11]. Weekly cross-correlation was found to correlate with self-reported habits in 1 study [11].

Strengths and Limitations

First, the 2 objective habit metrics used in this study are measures of consistency in timing, day-to-day, or week-to-week. However, not all habits translate into consistency in timing, and consistency in timing can originate from other factors than habit [9]. In such cases, the objective habit metrics will not reliably quantify the presence or absence of habit. On the other hand, the objective consistency metrics have the advantage that they do not endure biases associated with self-reported habit metrics, such as recall bias.

Second, the dataset used in this paper was very large, so the findings are likely to be generalizable to the whole population of participants enrolled in clinical trials. On the other hand, no conclusion about a specific disease, population, or time can be drawn, because there would be a large confounding influence of the underlying study design and specific medication characteristics. The main result is that, in general, adherence is related to objective habit.

Third, in this study, habit and medication adherence were computed over a participant's whole follow-up, until treatment discontinuation. However, for a single participant, habit strength may have changed throughout the duration of this study, for instance, because of the introduction of a new device for medication adherence monitoring. The evidence on the effects that electronic adherence monitoring causes on adherence itself is unclear [22,31]. The effects of electronic adherence monitoring on medication-taking habit have never been studied, to the best of the authors' knowledge.

Another reason that may have caused individual changes in habit strength is that some of the studies analyzed were testing habit-building interventions. In such settings, the objective habit metrics used in this paper can be computed over shorter periods to capture the dynamics of habit formation and maintenance. For instance, in the case of a specific habit-building intervention, Pironet et al [32] observed that habit strength increased after a 6-month intervention and remained stable 6 months after the intervention was stopped.

Future Directions

If a patient's implementation adherence is measured electronically, their habit can be computed continuously from the recorded data and serve as a support for an intervention, in-person or through a mobile app. The intervention could focus on providing support to build a better medication-taking habit. An example of such a mobile intervention is presented in the study by Stawarz et al [33].

This work investigated the relationship between habit and the implementation component of medication adherence. Other, evenly important questions are whether objective habit indices predict the quality of implementation over the longer term, and if objective habit indices predict early treatment discontinuation.

Conclusion

In this study, the correlation between objective habit metrics, computed from electronic medication intake data, and objective medication adherence was assessed in a large database of participants enrolled in past clinical trials. The 2 objective habit metrics tested in this work were correlated to the 2 objective adherence metrics. These 4 pairwise correlations all imply that a lower variability in the pattern of medication intakes, be it day-to-day or week-to-week, is correlated with higher medication adherence. However, no variability is not the norm.

Objective habit metrics allow us to better identify patients who might benefit from habit-building support. If a patient uses electronic monitoring, the habit metrics can even be automated and computed in real time, directly reflecting habit changes and allowing for timely intervention. In addition, some effective interventions for medication adherence are based on developing or improving habits [5]. In such settings, electronic monitoring can be used to assess the effect of the habit-building intervention in real time [32].

XSL•FO
RenderX

Acknowledgments

The authors wish to thank Eric Tousset, MSc, from AARDEX Group for his help with the statistical analyses. No generative artificial intelligence model was involved in any portion of this work.

Data Availability

The individual values of objective medication adherence (proportion of correct days and proportion of doses taken) and objective habit (SD of the hour of intake and weekly cross-correlation) can be downloaded on AARDEX Group's website, in the section "Knowledge Center."

Authors' Contributions

AP handled the conceptualization, data curation, formal analysis, methodology, visualization, and writing of the original draft. BV worked on the conceptualization, methodology, supervision, visualization, and review and editing of the writing. LAP did the conceptualization, methodology, supervision, and review and editing of the writing.

Conflicts of Interest

AP is an employee of AARDEX Group. BV is a shareholder of AARDEX Group. AARDEX Group commercializes the MEMS, the hardware and software that was used to record the electronic medication adherence data used in this study.

Multimedia Appendix 1
Objective habit strength metrics.
[DOCX File , 18 KB - ijmr_v14i1e63987_app1.docx]

References

1. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, ABC Project Team. A new taxonomy for describing and defining adherence to medications. Br J Clin Pharmacol 2012;73(5):691-705 [FREE Full text] [doi: 10.1111/j.1365-2125.2012.04167.x] [Medline: 22486599]
2. Khan R, Socha-Dietrich K. Investing in medication adherence improves health outcomes and health system efficiency: adherence to medicines for diabetes, hypertension, and hyperlipidaemia. In: OECD Health Working Papers. Paris: OECD Publishing; 2018:105.
3. Jimmy B, Jose J. Patient medication adherence: measures in daily practice. Oman Med J 2011;26(3):155-159 [FREE Full text] [doi: 10.5001/omj.2011.38] [Medline: 22043406]
4. Phillips LA, Leventhal H, Leventhal EA. Assessing theoretical predictors of long-term medication adherence: patients' treatment-related beliefs, experiential feedback and habit development. Psychol Health 2013;28(10):1135-1151. [doi: 10.1080/08870446.2013.793798] [Medline: 23627524]
5. Russell CL, Hathaway D, Remy LM, Aholt D, Clark D, Miller C, et al. Improving medication adherence and outcomes in adult kidney transplant patients using a personal systems approach: SystemCHANGE™ results of the MAGIC randomized clinical trial. Am J Transplant 2020;20(1):125-136 [FREE Full text] [doi: 10.1111/ajt.15528] [Medline: 31291507]
6. Costa E, Giardini A, Savin M, Menditto E, Lehane E, Laosa O, et al. Interventional tools to improve medication adherence: review of literature. Patient Prefer Adherence 2015;9:1303-1314 [FREE Full text] [doi: 10.2147/PPA.S87551] [Medline: 26396502]
7. Kini V, Ho PM. Interventions to improve medication adherence: a review. JAMA 2018;320(23):2461-2473. [doi: 10.1001/jama.2018.19271] [Medline: 30561486]
8. Wood W, Quinn JM, Kashy DA. Habits in everyday life: thought, emotion, and action. J Pers Soc Psychol 2002;83(6):1281-1297. [Medline: 12500811]
9. Volpp KG, Loewenstein G. What is a habit? Diverse mechanisms that can produce sustained behavior change. Organ Behav Hum Decis Process 2020;161:36-38. [doi: 10.1016/j.obhdp.2020.10.002]
10. Rajpura JR. Capsule commentary on Brooks et al., strategies used by older adults with asthma for adherence to inhaled corticosteroids. J Gen Intern Med 2014;29(11):1531 [FREE Full text] [doi: 10.1007/s11606-014-2962-2] [Medline: 25150031]
11. Phillips LA, Pironet A, Vrijens B. Evaluating objective metrics of habit strength for taking medications. J Behav Med 2023;46(4):632-641. [doi: 10.1007/s10865-023-00392-z] [Medline: 36662351]
12. 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;22(4):e17883 [FREE Full text] [doi: 10.2196/17883] [Medline: 32343250]

XSL•FO
RenderX

13. Hoo ZH, Wildman MJ, Campbell MJ, Walters SJ, Gardner B. A pragmatic behavior-based habit index for adherence to nebulized treatments among adults with cystic fibrosis. Patient Prefer Adherence 2019;13:283-294 [FREE Full text] [doi: 10.2147/PPA.S186417] [Medline: 30863018]
14. Conn VS, Ruppar TM. Medication adherence outcomes of 771 intervention trials: systematic review and meta-analysis. Prev Med 2017;99:269-276 [FREE Full text] [doi: 10.1016/j.ypmed.2017.03.008] [Medline: 28315760]
15. Verplanken B, Orbell S. Reflections on past behavior: a self-report index of habit strength. J Appl Soc Pyschol 2003;33(6):1313-1330 [FREE Full text] [doi: 10.1111/j.1559-1816.2003.tb01951.x]
16. Gardner B, Abraham C, Lally P, de Bruijn G. Towards parsimony in habit measurement: testing the convergent and predictive validity of an automaticity subscale of the Self-Report Habit Index. Int J Behav Nutr Phys Act 2012;9:102 [FREE Full text] [doi: 10.1186/1479-5868-9-102] [Medline: 22935297]
17. Phillips LA, Burns E, Leventhal H. Time-of-day differences in treatment-related habit strength and adherence. Ann Behav Med 2021;55(3):280-285 [FREE Full text] [doi: 10.1093/abm/kaaa042] [Medline: 32542355]
18. van de Vijver I, Brinkhof LP, de Wit S. Age differences in routine formation: the role of automatization, motivation, and executive functions. Front Psychol 2023;14:1140366 [FREE Full text] [doi: 10.3389/fpsyg.2023.1140366] [Medline: 37484115]
19. Linnemayr S, Odiit M, Mukasa B, Ghai I, Stecher C. Incentives and reminders to improve long-term medication adherence (INMIND): impact of a pilot randomized controlled trial in a large HIV clinic in Uganda. J Int AIDS Soc 2024;27(6):e26306 [FREE Full text] [doi: 10.1002/jia2.26306] [Medline: 38923298]
20. Phillips LA, Cohen J, Burns E, Abrams J, Renninger S. Self-management of chronic illness: the role of 'habit' versus reflective factors in exercise and medication adherence. J Behav Med 2016;39(6):1076-1091. [doi: 10.1007/s10865-016-9732-z] [Medline: 26980098]
21. Russell CL, Conn VS, Ashbaugh C, Madsen R, Hayes K, Ross G. Medication adherence patterns in adult renal transplant recipients. Res Nurs Health 2006;29(6):521-532 [FREE Full text] [doi: 10.1002/nur.20149] [Medline: 17131276]
22. McGrady ME, Ramsey RR. Using electronic monitoring devices to assess medication adherence: a research methods framework. J Gen Intern Med 2020;35(9):2707-2714 [FREE Full text] [doi: 10.1007/s11606-020-05905-z] [Medline: 32440997]
23. Stecher C, Mukasa B, Linnemayr S. Uncovering a behavioral strategy for establishing new habits: evidence from incentives for medication adherence in Uganda. J Health Econ 2021;77:102443 [FREE Full text] [doi: 10.1016/j.jhealeco.2021.102443] [Medline: 33831632]
24. Exemptions (2018 requirements). Office for Human Research Protections. 2019. URL: https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/common-rule-subpart-a-46104/index.html [accessed 2025-01-24]
25. Altman DG. Practical Statistics for Medical Research. Boca Raton, FL: CRC Press; 1999:293-295.
26. Cohen J, Cohen P, West S, Aiken L. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Mahwah, NJ: Lawrence Erlbaum Associates; 2003:49-50.
27. Myers L, Sirois MJ. Spearman correlation coefficients, differences between. In: Encyclopedia of Statistical Sciences. Hoboken, NJ: Wiley-Interscience; 2006:7901-7903.
28. Van Rossum G, Drake FL. Introduction to Python 3: Python Documentation Manual Part 1. Scotts Valley, CA: CreateSpace; 2009.
29. Kardas P, Lewek P, Matyjaszczyk M. Determinants of patient adherence: a review of systematic reviews. Front Pharmacol 2013;4:91 [FREE Full text] [doi: 10.3389/fphar.2013.00091] [Medline: 23898295]
30. Liu X, Lewis JJ, Zhang H, Lu W, Zhang S, Zheng G, et al. Effectiveness of electronic reminders to improve medication adherence in tuberculosis patients: a cluster-randomised trial. PLoS Med 2015;12(9):e1001876 [FREE Full text] [doi: 10.1371/journal.pmed.1001876] [Medline: 26372470]
31. Acosta FJ, Ramallo-Fariña Y, Bosch E, Mayans T, Rodríguez CJ, Caravaca A. Antipsychotic treatment dosing profile in patients with schizophrenia evaluated with electronic monitoring (MEMS®). Schizophr Res 2013;146(1-3):196-200. [doi: 10.1016/j.schres.2013.02.016] [Medline: 23474024]
32. Pironet A, Bartlett Ellis R, Stephen MB, Yerram P, Wakefield M, Awopetu D, et al. The SystemCHANGE intervention improves medication-taking habit. Germany: Springer; 2024 Presented at: Proceedings of the 27th annual meeting of ESPACOMP, the International Society for Medication Adherence; 2023 November 30-December 1; Budapest, Hungary p. 328-358. [doi: 10.1007/s11096-023-01688-5]
33. Stawarz K, Cox A, Blandford A. Don't forget your pill!: designing effective medication reminder apps that support users' daily routines. : ACM; 2014 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2014 April 26-May 1; Toronto Ontario Canada p. 2269-2278. [doi: 10.1145/2556288.2557079]

Abbreviations

MEMS: Medication Event Monitoring System

XSL•FO
RenderX

Edited by T de Azevedo Cardoso; submitted 05.07.24; peer-reviewed by B Gardner, I Ghai, S Linnemayr; comments to author 12.09.24; revised version received 04.11.24; accepted 08.01.25; published 06.02.25.

Please cite as:
Pironet A, Phillips LA, Vrijens B
Correlation Between Objective Habit Metrics and Objective Medication Adherence: Retrospective Study of 15,818 Participants From Clinical Studies
Interact J Med Res 2025;14:e63987
URL: https://www.i-jmr.org/2025/1/e63987
doi:10.2196/63987
PMID:

©Antoine Pironet, L Alison Phillips, Bernard Vrijens. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 06.02.2025. 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 Fortified Malt-Based Food on Immunity Outcomes in School Children in India: Cluster Randomized Controlled Trial

Anuradha Khadilkar[1], MD, DCH; Vinay Rawat[2], MBBS, MD; Jaladhi Bhatt[3], BSc, MSc; Devyani Chaturvedi[4], MSc; Vivek Garg[2], MBBS, MD; Pankaj Verma[2], MBBS, DNB, MSc

[1]Jehangir Clinical Development Centre, Jehangir Hospital Premises, Bund Garden Road, Sangamvadi, Pune, India

[2]GlaxoSmithKline Consumer Healthcare Limited, Gurgaon, India

[3]Hindustan Unilever Limited, Gurgaon, India

[4]Unilever USA, Newark, CA, United States

Corresponding Author:
Anuradha Khadilkar, MD, DCH
Jehangir Clinical Development Centre, Jehangir Hospital Premises, Bund Garden Road, Sangamvadi, Pune, India

Abstract

Background: Nutritional inadequacy and consequent diminished immunity among school-age children is a public health problem in India. Nutrition interventional studies using a cluster randomized controlled trial (RCT) design can avoid ethical issues inherent in double-blind individual RCTs in children involving daily administration of an empty-calorie placebo.

Objective: We tested the hypothesis that daily administration of a fortified malt-based food (FMBF), a multinutrient supplement, would improve immunity outcomes against common infectious diseases, nutritional status, and gut health in Indian school-age children by using a cluster RCT design. This report presents the study design attributes and the baseline characteristics of the study population.

Methods: This was an open-label, 2-arm, parallel-group, matched-pair cluster RCT, stratified by gender, in children aged ≥7 to ≤10 years old with height-for-age *z* scores (HAZ) of ≥−3 to ≤−1 and good general health. Four schools located in Pune city in India participated in the study. Each school was deemed as a cluster and was randomized to the test group (FMBF and dietary counseling) or control group (dietary counseling alone). A total of 924 participants from the 4 randomized schools were enrolled in the study.

Results: Observed mean age (SD) was 8.0 (SD 0.81; range: 7-10) years. There was no significant difference in mean age (*P*=.06), gender (*P*=.55), race (*P*>.99), HAZ category (*P*=.051), HAZ (*P*=.17), and BMI (*P*=.03). A very large proportion of children had micronutrient inadequacies in terms of vitamin D (97.5%), folate (79.2%), zinc (66%), and vitamin A (34.3%) at baseline. The study design meant that administration of the study intervention at a cluster level was easy. Mean compliance with the test product was 99.99% and retention in the study was 98%.

Conclusions: The findings highlight the extent of nutritional inadequacies in Indian school-age children, reaffirming the need for nutritional strategies to optimize the nutritional status among these children. A cluster RCT design can be effectively used in nutritional intervention trials with children by maintaining high compliance and retention.

Trial Registration: ClinicalTrials.gov NCT02542865; https://clinicaltrials.gov/study/NCT02542865

(Interact J Med Res 2025;14:e54189) doi:10.2196/54189

KEYWORDS

multinutrient supplement; study design; Indian children; immunity; fortified malt-based food; school-based

Introduction

Nutritional inadequacy has been the major and long-standing cause of nutritional and health concerns and developmental shortcomings in Indian children [1]. Despite numerous health care advances, the burden of major communicable diseases in children continues to be high in India compared to other countries [2,3]. As per the 2021 Global Burden of Disease study, diarrheal diseases and respiratory infections among children (5 - 14 y) were 2 of the 5 leading infectious causes of disability-adjusted life-years in the country [3]. Furthermore, child and maternal malnutrition (nutritional deficiencies, diarrhea, lower respiratory infections, and other common infectious diseases) were among the top 5 leading risk factors for disability-adjusted life-years [3].

In terms of nutritional status, approximately 60% of the Indian pediatric population has protein-energy malnutrition [4-8]. A recent nationwide survey further reported a deficiency in zinc, iron, and folate, as well as vitamins A, B12, and D in

XSL•FO
RenderX

14% - 32% of preschool children (aged 1-4 y), 17% - 28% of school-age children (aged 5 - 9 y), and 16% - 37% of adolescents (aged 10 - 19 y) [9].

Nutritional inadequacy hampers the overall development of children as it adversely affects their immunity, normal physiological functions, physical growth, and cognitive development [10-14]. Such inadequacies often coexist with infectious diseases and exhibit complex interactions, leading to a vicious cycle of malnutrition and infections [15-18]. This is evidenced by the fact that child undernutrition is responsible for 22% of the country's burden of disease [19].

As per the recent Global Nutrition Report, India performs poorly across standard child nutritional measures, accounting for nearly 39% of stunting and 15% of wasting worldwide [20]. On the national level for children under 5 years of age, data from the 2015 - 2016 National Family Health Survey showed a reduced prevalence of child stunting and underweight (38% and 36%) since 1992 - 1993 (52% and 53%) [21], however, it is still alarmingly high considering the multisectoral nutritional policies and initiatives that have been implemented [22]. Although there is sufficient anthropometry-related nutritional data for the under-5 age group, there is a dearth of national-level anthropometric data for school-age children. The Comprehensive National Nutrition Survey (CNNS) was the first national survey to measure anthropometry in children aged 5 to 14 years. According to the CNNS 2016 - 2019, the prevalence of stunting and underweight for children in the age group of 5 - 9 years was nearly 22% and 35%, respectively. However, the prevalence of stunting and underweight varied with school status, with a higher percentage of both among children who were out of school compared to school-going children (stunting: 38% vs 20%; underweight: 45% vs 34%) [9].

Evidently, at the individual level, appropriate nutrition actions in children can help bridge the nutritional gap, reduce the morbidity burden through improved immunity, and contribute to their overall health during the crucial formative and learning years [13,23-31]. Nutritional supplementation is one of the widely accepted approaches out of several effective nutritional strategies available worldwide [32,33]. Of note, in the Indian scenario, nutritional supplements have mainly been studied for their effect on the nutritional status and immunity of children <5 years of age. However, since school-age children represent about one-fifth of the total Indian population [34], understanding the effect of nutritional supplements in this cohort is equally essential. In addition, of the existing literature, none of the studies have comprehensively evaluated any nutritional supplement for its overall effect on the growth and immunity/health outcomes of Indian school-age children [35-38].

Thus, we conducted a study to provide a comprehensive overview of one such supplement, a multinutrient fortified malt-based food (FMBF), for its effect on immunity as well as other health aspects, in combination with dietary counseling, in a large group (>900) of Indian school-age children (>7 and <10 y). This paper gives an overview of the study design and baseline characteristics of the study population.

Methods

Study Design

This was a single-center, multiple-site, open-label, parallel-group, school-based, matched-pair cluster randomized controlled trial (RCT) conducted with Indian school-age children aged ≥7 to ≤10 years (ClinicalTrials.gov registration number NCT02542865). The study was conducted over 10 visits that consisted of screening (visit 1), baseline assessment (visit 2), dietary counseling (visit 3 to 9), and an end-of-study visit (visit 10). Nutritional intervention with an FMBF for the test group was done every day from visit 3 to 9 over a period of 9 months.

After obtaining written permission from management boards, 4 nongovernment schools from Pune, India, with a coeducation system and no history of outbreaks/clusters of infectious disease cases in the prior month consented to participate in the study. Schools were matched for size (with respect to the number of children aged 7 to 10 y), socioeconomic profile of the students, and type of school (government-subsidized or private) and 2 matched pairs were formed. Each school was deemed a cluster and assigned a randomization number in ascending numerical order. A randomization schedule with details on treatment assignment was prepared by an independent statistician from the Biostatistics Department, GlaxoSmithKline Consumer Healthcare, using validated internal software. The researchers involved in the study did not know the details of the schedule. The schedule was stored and concealed in an internal electronic document management system. Based on the randomization schedule and school list, each school within the matched pairs was randomized to receive either the multinutrient FMBF along with dietary counseling (test group) or dietary counseling alone (control group) during the intervention phase. The study assessed the effect of the FMBF on the overall development of these school-age children (immunity, nutritional status, gut health, anthropometry, dietary status), along with the FMBF's safety profile.

For quality control, all equipment and instruments were regularly calibrated, an accredited lab was used, and tools were validated. Staff were trained and retrained at regular intervals. To maintain the quality standards of data collection as per protocol and regulatory standards, data monitoring and data collection and analysis were done by a contract research organization. In addition, the study was audited by the Clinical Data Quality Assurance team at GlaxoSmithKline US and received an audit certificate.

Ethical Considerations

The study was conducted in accordance with the International Council on Harmonization Guidelines for Good Clinical Practice and the Declaration of Helsinki. The protocol and all the amendments were approved by Ethics Committee Jehangir Clinical Development Centre Pvt Ltd (ECR/352/Inst/MW2013) prior to study initiation. Written informed consent was obtained from one of the parents or a legally acceptable representative (LAR) of each participant and written assent was obtained from all participants before any study-specific procedures were performed. Participants didnot receive any compensation for participating in the clinical study but all participants travel cost

XSL•FO
RenderX

for study visit was reimbursed as per approved amount by Ethics Committee and same was mentioned in Informed Consent. The data collected for clinical study was stored by protecting privacy and maintaining confidentiality.

Study Inclusion and Exclusion Criteria

The key study inclusion and exclusion criteria are summarized in Textbox 1.

Textbox 1. Key study eligibility criteria.

Inclusion criteria

- Children of either gender
- Children aged 7 to 10 years (inclusive) with height-for-age *z* scores (HAZ) between −3 and -1 (inclusive)
- Children with good general and mental health in the opinion of the investigator or a medically qualified designee, defined as: (1) no clinically significant/relevant abnormalities in medical history or upon physical examination and (2) absence of any condition that could affect the child's safety/well-being or their ability to understand the study aspects

Exclusion criteria

- Children in care
- Children with an allergy or intolerance to any ingredient of the fortified malt-based food
- Children with severe anemia (hemoglobin <8 g/dL)
- Children with a history of use of immunosuppressive therapy (eg, oral corticosteroids or chemotherapy) in the 6 months prior to the screening visit
- Children with a recent history (2 months) of serious infections, injuries, and/or surgeries
- Children with a current or relevant history of any serious, severe, or unstable physical or psychiatric illness or any medical disorder that could have made them unlikely to complete the study
- Children who have consumed nutritional supplements and/or health food drinks on a regular basis (≥3 times a week) in the last 3 months
- Children with any other condition that presented an undue risk to them from the study interventions or procedures, in the opinion of the investigator or medically qualified designee

Study Interventions

The FMBF was prepackaged in sachets, each containing 27 g of multinutrient beverage powder. The nutrition profile of this supplement is presented in Table S1 in Multimedia Appendix 1. Participants from the test group consumed 27 g of this powder mixed in 150 mL of lukewarm water twice per day for 9 months. To avoid stomach fullness as a deterrent against compliance, the interval between the 2 doses was maximized. One dose was administered as soon as the participants entered the school and the second dose was given just prior to school dismissal. When a child was absent from school (eg, due to holidays or sickness), parents/LARS administered the 2 doses at home (one in the morning, one in the evening).

Nutritional inadequacy and the resulting immunity outcomes could be influenced by dietetic interventions implemented to improve nutritional intake. Background diet is a major confounder and it is imperative to track dietary quality throughout the study. Thus, to keep the background diets comparable between the 2 groups, the test group was administered the test product, while both the test and control group received dietary counseling. All study participants, their teachers, and their parents/LARs received age-specific dietary counseling to optimize the nutritional intake of the participants, which was based on the dietary guidelines for Indians established by the Indian Council of Medical Research–National Institute of Nutrition [39]. A total of 7 separate sessions of dietary counseling were administered to study participants and parents/LARs in both groups. The first 2 sessions involved the delivery of content and were mandatory to attend. The remaining 5 sessions were for follow-up, reinforcement, and problem-solving of content delivered in the first 2 sessions and these sessions were not mandatory to attend. The quality, content, and duration of the dietary counseling were equivalent in both the study groups.

Study End Points

The primary efficacy end point assessed the total number of ill days due to gastrointestinal (GI) and respiratory illness episodes in the participants over 9 months. GI illness was defined as an acute illness that included any of following symptoms: 3 or more loose, liquid, or watery stools or any vomiting in 24 hours [40]. Diarrhea or vomiting arising from noninfectious causes, such as irritable bowel syndrome or medications, were excluded. Respiratory illness was defined as an acute illness that included ≥1 of the following symptoms: runny nose, stuffy or blocked nose, cough, fever or chills, sore throat, or sneezing [41]. Symptoms arising due to noninfectious causes (eg, allergy) were excluded from the diagnosis. The secondary efficacy end points evaluated over 9 months included assessment of frequency and severity of the GI and respiratory illness episodes, school absenteeism due to these episodes, change from baseline in serum micronutrient levels (vitamins A, B12, D, and E, folate, and trace elements zinc, iron, copper, and selenium), change from baseline in macronutrient status, change from baseline in dietary diversity assessed using 24-hour individual dietary diversity score (IDDS), change from baseline in gut integrity as measured by a lactulose:mannitol test and urinary neopterin test, and change from baseline in BMI. Details of the evaluation

XSL•FO
RenderX

of secondary end points are presented in Table 1. Safety assessments included an evaluation of adverse events (AEs), treatment-emergent AEs, and serious AEs throughout the study. The laboratory assessments, vital signs, physical examinations, and hemoglobin measurements in these participants were conducted at baseline and at the end of the study.

XSL•FO
RenderX

Table . Secondary end point evaluation methods and tools.

Secondary efficacy end point	Frequency of assessment	Evaluation tool	Evaluation/test method
Frequency (number of episodes) of gastrointestinal and respiratory illnesses	Weekly basis during scheduled SPWR[a] visits	Diagnosis form	Calculated as total number of gastrointestinal and respiratory illness episodes, divided by the duration of the intervention, where each episode is defined as each instance of illness followed by at least 3 symptom-free days [40,42,43].
Severity of an illness episode	Weekly basis during scheduled SPWR visits	Diagnosis form	Assessed as per severity grading of mild, moderate, and severe as outlined, evaluated, and classified by a study physician [44].
School absenteeism due to gastrointestinal and respiratory illnesses	Weekly basis during scheduled SPWR visits	Diagnosis form	Calculated as number of days when a child failed to attend school because of gastrointestinal and/or respiratory illnesses.
Change from baseline in BMI	Twice in the study (at screening and at the end of the study visit)	• Portable stadiometer (SECA 213) for measuring a participant's height. • Standardized weighing scale (SECA 874) for measuring a participant's weight.	Calculated using the formula BMI=weight [kg]/(height[m])^2. All anthropometric measurements were performed using the guidelines adopted at the National Institute of Health–sponsored Arlie Conference [45].
Change from baseline in gut integrity/health	Twice in the study (at baseline and at the end of the study visit)	Urine testing	Urine lactulose:mannitol assessment by HPLC[b]. Urinary neopterin assessment by ELISA[c].
Change from baseline in mucosal immunity	Twice in the study (at baseline and at the end of the study visit)	Saliva testing	Salivary IgA[d] assessment by ELISA
Change from baseline in levels of the micronutrients vitamins A, B12, D (25-hydroxycholecalciferol), and E, folate, and trace elements selenium, zinc, copper, and iron	Twice in the study (at baseline and at the end of the study visit)	Blood testing	• Serum vitamin A, B12, D, and E, and folate assessment by ELISA • Serum copper assessment by 3, 5-Dibromo-PAESA • Serum zinc assessment by 5-Br-PAPS • Serum iron assessment by TPTZ[e] • Serum ferritin assessment by turbidimetry • Serum selenium assessment by ICP[f]
Change from baseline levels of ferritin, sTfR[g], CRP[h], and AGP[i]	Twice in the study (at baseline and at the end of the study visit)	Blood testing	• Serum sTfR and CRP assessment by immunoturbidimetry • Serum AGP assessment by turbidimetry
Change from baseline in dietary diversity score	Twice in the study (at baseline and at the end of the study visit)	• 24-hour dietary recall survey • Dietary diversity questionnaire [46]	Based on data of the foods and beverages consumed in the last 24 hours as captured by a 24-hour dietary survey, the appropriate food groups in the questionnaire were selected. Dietary diversity scores were calculated by adding the number of food groups consumed by the child over the 24-hour recall period.
Change from baseline in energy, protein, carbohydrates, and fat consumption	Twice in the study (at baseline and at the end of the study visit)	24-hour dietary recall survey [47]	24-hour dietary recall survey

[a]SPWR: study physician weekly review.

XSL•FO
RenderX

[b]HPLC: high-performance liquid chromatography.
[c]ELISA: enzyme-linked immunosorbent assay.
[d]IgA: immunoglobulin A.
[e]TPTZ: 2,4,6-Tripyridyl-S-triazine.
[f]ICP: inductively coupled plasma.
[g]sTfR: soluble transferrin receptor.
[h]CRP: c-reactive protein.
[i]AGP: alpha-1-acid glycoprotein.

End Point Assessments

Study end point assessments were performed as per the study schedule presented in Table 2.

XSL•FO
RenderX

Table . Study baseline and end point assessment schedule.[a]

Activity	Visit 1[b]	Visit 2[c]	Dietary counseling visits[d] (visits 3-9)[e]	Visit 10[f]
Informed consent and assent[g]	✓			
Demographics	✓			
Medical history	✓			
Current/concomitant medication	✓	✓	✓	✓
General physical examination	✓			✓
Vital signs	✓			✓
Hemoglobin assessment using Pronto	✓			
Anthropometric measurements (height, weight, BMI, and height-for-age *z* scores)	✓			✓
Inclusion/exclusion criteria evaluation	✓			
Participant eligibility	✓			
Continued eligibility criteria		✓	✓	✓
24-hour dietary recall and dietary diversity survey	✓			✓
Parental/LAR training to complete “symptom checklist”[h]	✓			
Dispense blank parent/LAR symptom checklist and product compliance report forms[i]	✓			
Sample collection for urinary neopterin test and lactulose:mannitol test		✓		✓
Sample collection for analysis of serum ferritin, serum transferrin receptor, c-reactive protein, alpha-1-acid glycoprotein, salivary immunoglobulin A[j], and nutritional biochemistry		✓		✓
Dietary counseling			✓	
Product administration (applicable to the test group only)[k]			✓	
Product compliance check and collecting empty product sachets (applicable to test group only)			✓	
Adverse event monitoring		✓	✓	✓
Study completion and medical sign off				✓

[a]Study Physician Weekly Review was initiated 1 week (+ up to 2 days) post first dietary counselling visit of the participant and was continued on a weekly basis (+ up to 2 days) until end of study visit.

XSL•FO
RenderX

[b]Visit 1: screening visit.

[c]Visit 2: baseline visit (+1 to 21 days post–screening visit).

[d]Sessions number 3, 4, 5, 6, and 7 could be postponed in view of long school holidays and exams. Session number 7 was not to be conducted any later than the last day of month 8 of the intervention period. A minimum of a 4-week gap was maintained between sessions number 2, 3, 4, 5, 6, and 7.

[e]Visit 3/session 1: day 1 to 3 of week 1 from baseline visit; visit 4/session 2: week 2 from baseline visit and up to 3 days; visit 5/session 3: week 6 from baseline visit and up to 3 days; visit 6/session 4: week 10 from baseline visit and up to 3 days; visit 7/session 5: week 14 from baseline visit and up to 3 days; visit 8/session 6: week 18 from baseline visit and up to 3 days; visit 9/session 7: week 22 from baseline visit and up to 3 days.

[f]Visit 10: end of study visit (+1 to 7 days after the end of 9 months from the baseline visit).

[g]Obtaining informed consent and assent and performing screening activities could be on different dates but informed consent and assent were obtained prior to performing and applying any screening criteria or procedures. Screening activities were done within 60 days of obtaining consent and assent.

[h]Could be done any time after obtaining consent and assent but prior to baseline visit.

[i]Parent/legally acceptable representative symptom checklist and product compliance report forms were collected and reviewed for each weekly review with the physician.

[j]Saliva samples were collected for the assessment of salivary immunoglobulin A at visit 2 and visit 10.

[k]Intervention began +1 to 3 days from the baseline visit and continued for 9 months. Compliance checks (test product empty sachets were collected) and AEs were recorded throughout the intervention period.

To evaluate primary and secondary immunity outcomes, the parents/LARs of the participants regularly updated the symptom checklist and compliance report forms for study interventions, which were collected and analyzed during each study physician weekly review (SPWR) throughout the study. During the SPWR, the start/end date of GI and respiratory illness episodes along with the frequency, severity (mild, moderate, severe), and school absenteeism due to these episodes were noted in the diagnosis form by the study physicians based on details obtained from the following:

- Symptom checklist completed by parents/LARs
- External clinical record if treated by an external physician
- Direct information from participants or parents/LARs
- Clinical history and examination by study physician

Details of the evaluation methods and tools used for assessment of secondary immunity outcome end points, anthropometry, serum biochemistry, and dietary assessment are provided in Table 1.

During the study, the AEs and serious AEs were monitored in addition to those reported by participants or parents/LARs to the study physicians as and when observed or through the weekly symptom checklists. The AEs were regarded as treatment-emergent AEs if they occurred on or after the first administration of the FMBF in the test group or on or after the first dietary counseling session in the control group.

A compliance check of the study interventions was conducted as per the schedule presented in Table 2. Concomitant medications were recorded at each study visit. Except for immunosuppressive therapies (which were never used in the study), there were no restrictions on medications and medical treatments. However, the use of medications (eg, antibiotics, antipyretics) was standardized across all study sites.

A noninvasive, spot-check monitoring device (Masimo Corporation's Pronto Pulse CO-Oximeter) was used to screen study participants in school settings based on their hemoglobin (Hb). This device uses transcutaneous spectrophotometry to measure Hb concentration in the blood, and the result is comparable to invasive laboratory-based Hb estimations in children [48]. Anthropometric measurements included the assessment of height, weight, BMI, and HAZ. HAZ was obtained using the World Health Organization Anthroplus software [49].

Sample Size

The sample size was based on the primary efficacy end point, the total number of ill days due to GI and respiratory illnesses. Based on the data from the Global Burden of Disease by the World Health Organization [50], the Institute for Health Metrics and Evaluation [51], and the Census of India [34], approximately 7 ill days due to GI and respiratory illnesses per 9 months was assumed in this study population. Thus, the sample size was calculated based on the assumption of an average of 7 ill days in the control group with a difference of 10% (0.7 d) between the study groups.

No sample size considerations were taken into account for secondary objectives. The simulated data were also used based upon the assumption of a Poisson distribution per study treatment. On observation of these simulated data, it was considered that these Poisson distributions were not heavily skewed and, as a result, normal distributions could be assumed. Consequently, a normality assumption was adopted for both the generation of sample size and the intended approach to the statistical analysis. Because of the assumption of Poisson distribution, estimates of variability were calculated as the square root of the mean. The mean of the control was assumed to be 7.0 (SD 2.65), while the mean of the test group was 6.3 (SD 2.51). In order to achieve 80% power, 215 participants per treatment arm (total=430) were required to complete the study in case of individual randomization. This assumed a 5% level of significance (2-tailed *t* test). As this study used cluster randomization, to retain equivalent power to an individually randomized trial, design effect (DE), which is a function of the intracluster correlation coefficient (ICC), was considered for sample size calculations. The scientific literature suggested that DE due to a cluster sampling strategy should be assumed to be 1.5 [52]. A similar assumption was made in a study that used 3-stage systematic cluster sampling [53]. The following formula was used [54]:

$$N* = N \times DE$$

XSL•FO
RenderX

Where N* is the number of participants required for a cluster RCT, N is the number of participants required for an individual RCT, and DE is the design effect.

Since DE was assumed to be 1.5, the number of evaluable participants required to complete the study was determined to be 323 per arm (total=646). To allow for a 30% dropout rate, a total of 924 participants were to be randomized. Approximately 1300 participants were to be screened to randomize approximately 924 participants in order to obtain 646 participants completing the study.

Statistical Analysis

Analysis Population

Safety population was defined as participants who received ≥1 dose of the study interventions (only dietary counseling [control group] or both FMBF and dietary counseling [test group]) during the study. The intention-to-treat (ITT) population comprised all participants in the safety population with any postinterventional assessment (ie, SPWR), according to the study group assigned. The modified intention-to-treat (mITT) population was defined as all participants in the safety population with any postinterventional assessment (SPWR) who completed the entire intervention phase and attended the end-of-study visit.

Analysis Methods

Access to the treatment code was appropriately restricted during the process of blinded data review involving clinical research scientists and statisticians. The database with potentially unblinded information was separately reviewed for study population inclusion/exclusion by unblinded clinical research scientists. The database was later validated by a blinded reviewer prior to its release for analysis.

Analysis of the primary efficacy end point was done on the mITT population and the analysis of secondary and exploratory efficacy end points was done on both the mITT and ITT population. They were analyzed using the analysis of covariance (ANCOVA) model (SAS Studio version 9.4 or higher; SAS Institute Inc). The ANCOVA had cluster (school) as a random effect; study group and gender as fixed effects; and baseline assessments, baseline serum levels, and baseline IDDS as covariates. Unlike other study end points, frequencies for severity (mild, moderate, severe) of GI/respiratory illness episodes as well as IDDS categories (low, moderate, high) were compared using the χ^2 test (if frequencies were >5%). If frequencies were ≤5% in any study group, then these were compared using the Fisher exact test. Adjusted means for each study group, 95% confidence intervals, within-group *P* values, difference between the study groups, 95% confidence intervals of the difference, and the between-group *P* values were also presented based on the above statistical model. All statistical tests of hypotheses were 2-sided and used a level of significance of α=.05.

Demographics, baseline characteristics, and safety assessment (including laboratory tests) summaries were produced for the safety, ITT, and mITT populations. Descriptive statistics including n (%), mean (SD), minimum, and maximum (as applicable) were provided. Baseline characteristics were compared between the treatment groups using an independent *t* test for continuous variables and χ^2 test for categorical variables. A 2-sided significance level of 5% was used for the comparisons.

The intervention was targeted at the cluster level; however, outcomes were measured at the individual level. Thus, all ANCOVA analyses included cluster as a random effect to introduce correlation between children from the same school, which was estimated by the ICC. The assumptions of normality and homogeneity of variance were investigated. Violations of these assumptions were overcome using suitable transformation (eg, log).

Results

Overview

The study was conducted between July 2017 and July 2018. Recruitment of all participants was completed in a 4-month period. After screening and baseline assessment, each recruited participant remained in the study for 9 months to receive the interventions and undergo assessments.

The mean product compliance in the test group was 99.99% (SD 0.02, minimum 99.69% and maximum 100%). Mean compliance with both mandatory and follow-up sessions of dietary counseling for participants as well as parents/LARs was 100% in both the test and control groups.

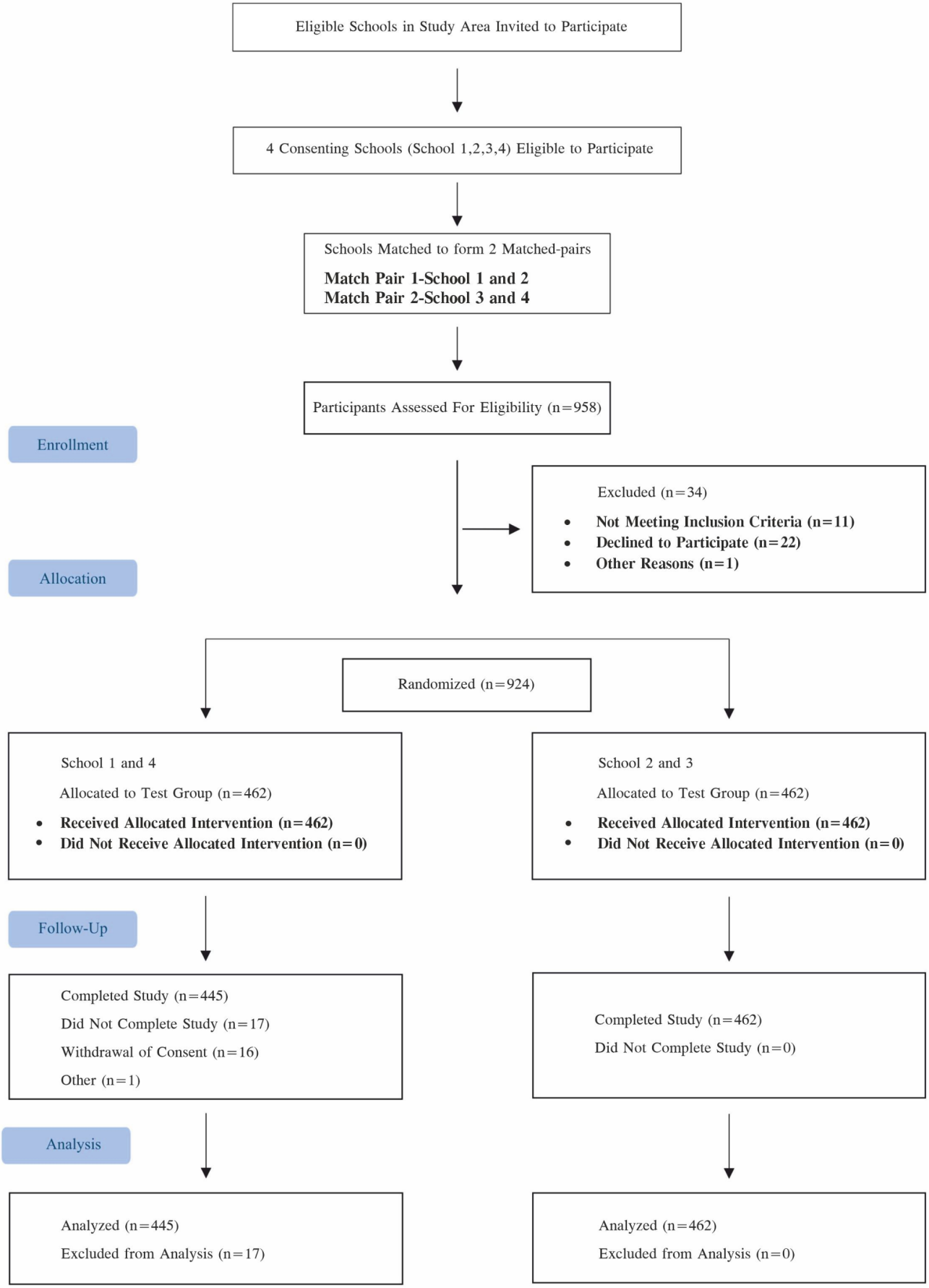
Demographic and Baseline Characteristics of Study Participants

Figure 1 depicts the study design and disposition of the study participants. Of 958 screened participants, 924 participants were randomized to the test group and the control group in a 1:1 proportion (n=462 each). A total of 907 participants (98.2%) completed the study (test: n=445; control: n=462), whereas 17 (3.7%) participants did not complete the study, either due to withdrawal of consent (test: n=16, 3.5%) or other unspecified reasons (test: n=1, 0.2%).

The demographic and baseline characteristics of the study participants are summarized in Table 3. All the participants had Central/South Asian heritage. Both the study groups (test vs control) were comparable in terms of mean age of the participants (8.0 y vs 7.9 y), gender distribution (male: 57.4% vs 55.4%), mean HAZ (−1.52 vs −1.48), and mean BMI (15.47 kg/m^2 vs 15.12 kg/m^2).

XSL•FO
RenderX

Figure 1. CONSORT flowchart of study design and participant disposition. CONSORT: Consolidated Standards of Reporting Trials.



XSL•FO
RenderX

Table . Demographic and baseline characteristics of study participants (safety population).

Parameters	Test group (n=460)	Control group (n=462)	Total (n=922)	*P* value
Age, years				
Mean (SD)	8.0 (0.78)	7.9 (0.84)	8.0 (0.81)	.06
Range	7-10	7-9	7-10	
Gender, n (%)				
Female	196 (42.6)	206 (44.6)	402 (43.6)	.55
Male	264 (57.4)	256 (55.4)	520 (56.4)	
Race, n (%)				
Central Asian/South Asian heritage	460 (100)	462 (100)	922 (100)	>.99
HAZ[a] category				
−3 to −2	74 (16.1)	54 (11.7)	128 (13.9)	.051
−2 to −1	385 (83.7)	408 (88.3)	793 (86)	
−1 to 0	1 (0.2)	0 (0)	1 (0.1)	
HAZ				
Mean (SD)	−1.52 (0.48)	−1.48 (0.41)	−1.50 (0.44)	.17
Range	−2.96 to −0.45	−3.00 to −1.00	−3.00 to −0.45	
BMI, kg/m^2				
Mean (SD)	15.47 (2.29)	15.12 (2.55)	15.29 (2.43)	.03
Range	9.39-23.68	10.14-27.09	9.39-27.09	

[a]HAZ: height-for-age *z* scores.

Baseline Nutrient Status of Study Participants

Tables 4 and 5 summarize the baseline micronutrient status of the study participants. The findings indicate a considerable burden of micronutrient inadequacy in these participants. The most frequently observed inadequacy was vitamin D (97.5%), followed by folate (79.2%), zinc (66%), and vitamin A (34.3%; Table 6). An overview of the 24-hour macronutrient intake status is depicted in Table 7. The nutritional intake status was similar between both the study groups (Tables 4 and 5).

Table . Summary of micronutrient status of the study participants at baseline (mITT[a] population).

Parameters	Reference range[b]	Test group (n=445), mean (SD)	Control group (n=462), mean (SD)	Total (n=907), mean (SD)	*P* value
Serum vitamin A	26.00 - 49.00 µg/dL	29.52 (6.30)	28.14 (5.41)	28.82 (5.89)	<.001
Serum vitamin B12	312.00 - 1237.00 pg/mL	374.38 (87.31)	375.67 (97.51)	375.04 (92.59)	.83
Serum vitamin D	30.00 - 100.00 ng/mL	17.48 (5.63)	17.89 (5.10)	17.69 (5.37)	.25
Serum vitamin E	0.30 - 0.90 mg/dL	0.49 (0.14)	0.47 (0.16)	0.48 (0.15)	.046
Serum folate	5.00 - 21.00 ng/mL	4.51 (0.73)	4.42 (0.71)	4.46 (0.72)	.06
Serum selenium	55.00 - 134.00 µg/L	83.63 (12.61)	85.91 (12.24)	84.79 (12.47)	.01
Serum zinc	78.00 - 105.00 µg/dL (aged 7 - 9 y), 78.00 - 118.00 µg/dL (females aged 10 y), 78.00 - 98.00 µg/dL (males aged 10 y)	72.17 (19.80)	72.58 (19.98)	72.38 (19.88)	.76
Serum copper	51.00 - 121.00 µg/dL	117.89 (15.92)	117.77 (16.08)	117.83 (15.99)	.91
Serum iron	50.00 - 120.00 µg/dL	70.09 (29.77)	71.43 (29.14)	70.77 (29.44)	.49

[a]mITT: modified intention-to-treat.

[b]As per the Intervein Lab, Ahmedabad, India, a National Accreditation Board for Testing and Calibration Laboratories–accredited lab.

XSL•FO
RenderX

Table . Summary of micronutrient inadequacy status in the study participants at baseline (safety population).

Parameter	Reference range	Status of micronutrient inadequacy, n (%)				*P* value
		Abnormality	Test group (n=460)	Control group (n=462)	Total (n=922)	
Serum vitamin A	26.00 - 49.00 µg/dL	Low	146 (31.7)	170 (36.8)	316 (34.3)	.11
Serum vitamin B12	312.00 - 1237.00 pg/mL	Low	121 (26.3)	148 (32)	269 (29.2)	.06
Serum vitamin D	30.00 - 100.00 ng/mL	Low	449 (97.6)	450 (97.4)	899 (97.5)	>.99
Serum vitamin E	0.30 - 0.90 mg/dL	Low	32 (7.0)	53 (11.5)	85 (9.2)	.02
Serum folate	5.00 - 21.00 ng/mL	Low	354 (77)	376 (81.4)	730 (79.2)	.10
Serum selenium	55.00 - 134.00 µg/L	Low	1 (0.2)	0 (0)	1 (0.1)	.50
Serum zinc	78.00 - 105.00 µg/dL (aged 7 - 9 y), 78.00 - 118.00 (females aged 10 y), 78.00 - 98.00 (males aged 10 y)	Low	307 (66.7)	302 (65.4)	609 (66)	.68
Serum copper	51.00 - 121.00 µg/dL	Low	0 (0)	1 (0.2)	1 (0.1)	>.99
Serum iron	50.00 - 120.00 µg/dL	Low	133 (28.9)	121 (26.2)	254 (27.6)	.38

Table . Prevalence of micronutrient inadequacy in Indian school-age children (current study vs literature findings).

Micronutrients	Prevalence of micronutrient inadequacies		
	Current study, %	Individual studies from literature, %	CNNS[a] 2019 [9], %
Serum vitamin A	34.3	9.8 - 43.9 [37,55,56]	22
Serum vitamin B12	29.2	7.4 - 67.2 [37,57,58]	17
Serum vitamin D	97.5	14.2 - 93.0 [37,56-62]	18
Serum vitamin E	9.2	No data found	Not reported
Serum folate	79.2	1.5 - 99 [37,57,63,64]	28
Serum selenium	0.1	No data found	Not reported
Serum zinc	66	0.7 - 57.1 [37,64,65]	17
Serum copper	0.1	No data found	Not reported
Serum iron (serum ferritin)	27.6	24.1 - 54.5 [58,64]	17

[a]CNNS: Comprehensive National Nutrition Survey.

XSL•FO
RenderX

Table . Summary of macronutrient status of the study participants at baseline (mITT [a] population).

Parameters	RDA[b]	Test group (n=445)		Control group (n=462)		Total participants (n=907)	*P* value
		Mean (SD)	% RDA	Mean (SD)	% RDA	Mean (SD)	
Energy (kcal/day)	1690	1395.14 (305.78)	82.5	1406.08 (352.75)	83.2	1400.71 (330.40)	.62
Protein (g/day)	29.5	38.84 (11.88)	131.6	42.41 (13.39)	143.8	40.66 (12.79)	<.001
Carbohydrates (g/day)	N/A[c]	189.48 (50.75)	N/A	202.39 (55.08)	N/A	196.05 (53.36)	<.001
Fats (g/day)	30	47.52 (16.57)	158.4	40.03 (16.41)	133.4	43.71 (16.89)	<.001

[a]mITT: modified intention-to-treat.

[b]RDA: recommended dietary allowance.

[c]N/A: not applicable.

Since nutritional status is usually influenced by gut health, the study participants were also evaluated for their gut integrity through the lactulose:mannitol test and urinary neopterin test. A total of 111 (12%) participants demonstrated an abnormal lactulose:mannitol ratio (reference level: <0.035; test: n=66, 14.3%; control: n=45, 9.7%), indicating malabsorption in these participants (Table 8). On the other hand, the urinary neopterin test was normal for all the study participants across both groups (reference range: 0.10 - 5.00 mmol/mol creatinine).

Table . Summary of baseline gut integrity/health using the lactulose:mannitol test and urinary neopterin test (mITT[a] population).

Parameter	Test group	Control group	*P* value
Lactulose mannitol test			.002
Mean (SD)	0.0504 (0.19891)	0.0218 (0.01150)	
Overall range min-max	0.000-2.200	0.000-0.132	
Urine neopterin test			.27
Mean (SD)	0.419 (0.3967)	0.388 (0.4410)	
Overall range min-max	0.10 - 2.78	0.10 - 4.22	

[a]mITT: modified intention-to-treat.

Discussion

Principal Findings

This study revealed concerning facts about nutritional and growth parameters in the apparently healthy school-age children. The biochemical statuses of many micronutrients, namely vitamin A (34%), vitamin B12 (29%), vitamin D (97%), folate (79%), zinc (66%), and iron (28%) were observed to be inadequate in the children. In this study, we also assessed the gut health of the study participants. There is a paucity of prevalence data on this parameter in apparently healthy children, particularly those without diarrhea. We observed an abnormal lactulose:mannitol ratio in a substantial percentage (111/922, 12%) of the study population. This indicates compromised intestinal permeability, thus affecting their nutrient absorption.

This study found high compliance with the test product (99.99%), with a high retention rate of 98% (907/924). Maintaining similar influence and motivation among children and parents, particularly in an open-label individual RCT, could have been very challenging. However, this being a cluster/school-level intervention, peer influence could have helped improve compliance and adherence. Strategical measures used in the study to improve retention included weekly compliance monitoring, regular meetings and briefings with parents at the school, random home visits, and telephonic follow-up with nonresponders or during out-of-station travels.

Statistical implications of this study's cluster RCT design necessitated inflating the sample size by 1.5 times the number needed for an individual RCT. Despite a relatively large sample size and a 9-month duration, the study was administratively more efficient because of the greater logistical convenience resulting from the operational ease of administration of the study intervention at the cluster level.

The magnitudes of nutritional inadequacies observed in this study were in line with most of the previously published literature (Table 8), reaffirming the presence of nutritional concerns in Indian school-age children [9,37,55-65]. However, when compared with CNNS estimates, the observed magnitudes of vitamin D, folate, and zinc deficiencies were relatively high. This may be due to higher cutoffs for serum vitamin D, folate, and zinc used in this study compared to the CNNS. Cutoffs in this study versus the CNNS are as follows: 30 ng/mL versus 12 ng/mL for vitamin D, 5 ng/mL versus 4 ng/mL for folate (equivalent to the cutoff of 151 ng/mL of red blood cell folate used in the CNNS), and 78 mcg/mL versus 65 mcg/mL for zinc. As a result, a greater number of children were included in the

XSL•FO
RenderX

nutrient-deficient group in this study. Moreover, due to sedentary and indoor lifestyles, vitamin D deficiency is known to occur more commonly among children in urban settings [66,67], as in this study conducted with participants residing in the urban setting of Pune city.

This was one of the largest nutritional intervention trials conducted with a matched-pair cluster RCT design on Indian school-age children. Cluster trials are conceptually able to improve acceptability and adherence to an assigned intervention, but these qualities are rarely assessed and reported in cluster trial publications [68,69]. In a long nutritional intervention trial, consumption fatigue or boredom can set in, resulting in poor compliance and subsequent attrition from the study. This study recorded high compliance with the test product, with a high retention rate. Maintaining similar influence and motivation among children and parents, particularly in an open-label individual RCT, could have been very challenging. However, this being a cluster/school-level intervention, peer influence could have helped improve compliance and adherence. Besides, strategic measures used in the study including weekly compliance monitoring; regular meetings and briefings with parents at the school; random home visits; and telephonic follow-up with nonresponders or during out-of-station travels had an impact on retention.

Limitations

The open-label design used in the study had great potential to pose a threat to the validity by introducing potential biases associated with such a design, but this was curtailed significantly by using specific and objective criteria for the diagnosing and recording of morbidities by physicians. Besides, results based on biochemistry were also objective by nature. Additionally, access to the treatment code was appropriately restricted during the process of blinded data review involving clinical research scientists and statisticians. The study involved the random allocation of clusters (ie, schools); however, participants' selection within a school was not random due to the cluster RCT design. Prior knowledge of treatment could still have led to recruitment bias, which cluster RCTs are usually vulnerable to [70]. However, the study did not seem to have been impacted by this bias, as shown by the high screening to enrollment ratio, with a low proportion of participants refusing consent during screening in both the control group (16/482, 1.3%) and test group (6/476, 3.3%).

Although study groups were matched for overall socioeconomic status, the extent to which individual participants were matched was not ascertained. There was always a chance that potential socioeconomic status differences and associated factors (eg, access to health care, use of medications, hygiene, and sanitation practices) had the potential to confound the results. This was minimized by harmonizing the treatment of all morbidities by similarly trained study physicians in schools as a part of adverse event management. It was known that the use of medications could impact study outcomes, particularly ill days, school absenteeism, and illness severity. Thus, the medications (eg, antibiotics, antipyretics) were standardized across all study sites, and were thus eventually similar across both study groups. Since there were different study physicians at all 4 sites, an interobserver bias cannot be excluded. This was mitigated by making objective study end points and the physician's discretion was used only while ruling out the allergic etiology of the illnesses. In addition, the training of the study physicians was harmonized with respect to quality, content, and duration.

Cluster randomization has statistical implications. The sample size was adjusted and planned to be increased 1.5 times compared to that needed for an individual RCT. In the primary analysis, the observed ICC, which measures the relatedness of clustered responses, was higher than anticipated, corresponding to a design effect more than 10 times higher than the protocol assumption. It is, however, comparable with estimates from the literature [71]. A higher number of clusters, with the same overall number of participants, could have increased statistical power. However, fewer clusters were used in the study due to practical reasons.

Conclusions

The study findings highlight the extent of nutritional inadequacies in urban school-age children in India, reaffirming the need for nutritional strategies to optimize the nutritional status among these children.

This study suggested that a cluster RCT design can be effectively used in nutritional intervention trials. The design mitigates the ethical issues inherent in double-blind individual RCTs in children involving daily administration of an empty-calorie placebo. The strategies used to achieve high recruitment and retention rates are valuable for the conduct of future studies in a similar population.

Acknowledgments

We would like to acknowledge the contribution of the Jehangir Clinical Development Centre (JCDC) team for the smooth conduct of and data collection for this clinical study. Authors VR, DC, VG, and PV worked at GSK Consumer Healthcare (now Hindustan Unilever Limited) during the clinical study conduct phase and are currently working as Independent Researcher, Delhi, India; Senior Nutrition Scientist at Smarty Pants, California, United States; Global Medical Affairs Manager, Alliance Pharmaceuticals, Chippenham, United Kingdom; and Advisor, Hindustan Unilever Limited, India, respectively. This research was funded by HUL Nutrition (formerly GSK Consumer Healthcare).

Authors' Contributions

VR, PV, and VG were mainly involved in the design of the study. AK was involved in investigation, methodology, and review of the manuscript. JB was involved in the planning, quality check, and supervising clinical study conduct. DC drafted the original

XSL•FO
RenderX

manuscript. The final version of the manuscript was carefully reviewed and approved by all authors. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

All authors except AK were employed by the funder (GSK Consumer Healthcare, now Hindustan Unilever Limited) while this clinical study was conducted. AK was the principal investigator for this clinical study and part of the clinical study center.

Multimedia Appendix 1
Supplementary table
[XLSX File, 15 KB - i-jmr_v14i1e54189_app1.xlsx]

Checklist 1
CONSORT checklist 2025. CONSORT: Consolidated Standards of Reporting Trials.
[DOCX File, 33 KB - i-jmr_v14i1e54189_app2.docx]

References

1. Singh A. Childhood malnutrition in India. In: Bhattacharya SK, editor. Perspective of Recent Advances in Acute Diarrhea: IntechOpen; 2020. [doi: 10.5772/intechopen.89701]
2. Jain L, Pradhan S, Aggarwal A, et al. Association of child growth failure indicators with household sanitation practices in India (1998-2021): spatiotemporal observational study. JMIR Public Health Surveill 2024 May 24;10:e41567. [doi: 10.2196/41567] [Medline: 38787607]
3. Global burden of disease study 2021. Institute for Health Metrics and Evaluation. URL: https://www.healthdata.org/research-analysis/library/global-burden-disease-2021-findings-gbd-2021-study [accessed 2025-06-17]
4. Krishnan M, Kalaiselvi K, Rajalakshmi P. A study of protein energy malnutrition in the school girls of a rural population. Int J Nutr Pharmacol Neurol Dis 2012;2(2):142. [doi: 10.4103/2231-0738.95985]
5. Midha T, Nigam S, S Martolia D, Kaur S. Prevalence and determinants of protein energy malnutrition among under five children in slums of Kannauj district. IJFCM 2020 Dec 28;5(4):218-222. [doi: 10.18231/2394-6776.2018.0050]
6. Tiwari S, Bandi J, Awasthi S, Sharma A. Assessment of prevalence of protein energy malnutrition in under 5 year children in an urban slum of Mumbai, India and to study associated factors. Int J Community Med Public Health 2017;3(5):1129-1134. [doi: 10.18203/2394-6040.ijcmph20161371]
7. Kumar T, Deswal BS. An epidemiological study of protein energy malnutrition among children below six years' age in an urban slum of Gurgaon, Haryana, India. Int J Community Med Public Health 2016;3(9):2431-2436. [doi: 10.18203/2394-6040.ijcmph20163051]
8. Shivaprakash NC, Joseph RB. Nutritional status of rural school-going children (6-12 years) of Mandya District, Karnataka. Int J Sci 2014;2(2):1-5 [FREE Full text]
9. Comprehensive national nutrition survey national report. : Ministry of Health and Family Welfare, Government of India, UNICEF and Population Council; 2019 URL: https://www.unicef.org/india/media/2646/file/CNNS-report.pdf [accessed 2025-06-17]
10. Investing in the future. A united call to action on vitamin and mineral deficiencies. : Micronutrient Initiative; 2009 URL: https://www.nutritionintl.org/wp-content/uploads/2019/03/Investing_in_the_future_Summary.pdf [accessed 2025-06-17]
11. Eichler K, Hess S, Twerenbold C, Sabatier M, Meier F, Wieser S. Health effects of micronutrient fortified dairy products and cereal food for children and adolescents: a systematic review. PLoS One 2019;14(1):e0210899. [doi: 10.1371/journal.pone.0210899] [Medline: 30673769]
12. Best C, Neufingerl N, Del Rosso JM, Transler C, van den Briel T, Osendarp S. Can multi-micronutrient food fortification improve the micronutrient status, growth, health, and cognition of schoolchildren? A systematic review. Nutr Rev 2011 Apr;69(4):186-204. [doi: 10.1111/j.1753-4887.2011.00378.x] [Medline: 21457264]
13. Shatrugna V, Balakrishna N, Krishnaswamy K. Effect of micronutrient supplement on health and nutritional status of schoolchildren: bone health and body composition. Nutrition 2006 Jan;22(1 Suppl):S33-S39. [doi: 10.1016/j.nut.2005.07.010] [Medline: 16426961]
14. Kuther T. Chapter 9 – lifespan development: lives in context. In: Physical and Cognitive Development in Middle Childhood 2020.
15. Bhaskaram P. Micronutrient malnutrition, infection, and immunity: an overview. Nutr Rev 2002 May;60(5 Pt 2):S40-S45. [doi: 10.1301/00296640260130722] [Medline: 12035857]
16. Kar K, Pradhan S, Samal BP. Morbidity pattern of school children: a study conducted in the urban field practice area of a tertiary health care centre of Odisha, India. Int J Res Med Sci 2018;6(7):2489. [doi: 10.18203/2320-6012.ijrms20182841]
17. Patel N, Gunjana G, Patel S, Thanvi R, Sathvara P, Joshi R. Nutrition and health status of school children in urban area of Ahmedabad, India: comparison with Indian Council of Medical Research and body mass index standards. J Nat Sci Biol Med 2015;6(2):372-377. [doi: 10.4103/0976-9668.160010] [Medline: 26283833]

XSL•FO
RenderX

18. Mandlik R, Chiplonkar S, Khadilkar V, et al. Prevalence of infections among 6-16 years old children attending a semi-rural school in Western Maharashtra, India. Indian J Child Health 2015;2:182-186. [doi: 10.32677/IJCH.2015.v02.i04.010]
19. Gragnolati M, Bredenkamp C, Shekar M, Das Gupta M, Lee YK. India's undernourished children: a call for reform and action. : The World Bank; 2006. [doi: 10.1596/978-0-8213-6587-8]
20. Global nutrition report 2016: from promise to impact: ending malnutrition by 2030. : International Food Policy Research Institute; 2016 URL: https://media.globalnutritionreport.org/documents/English_full_report.pdf [accessed 2025-06-17]
21. National family health survey (NFHS-4), 2015-16. : International Institute for Population Sciences and ICF; 2017 URL: https://dhsprogram.com/pubs/pdf/fr339/fr339.pdf [accessed 2025-06-17]
22. Sreedevi A. An overview of the development and status of national nutritional programs in India. J Med Nutr Nutraceut 2015;4(1):5. [doi: 10.4103/2278-019X.146146]
23. Schaible UE, Kaufmann SHE. Malnutrition and infection: complex mechanisms and global impacts. PLoS Med 2007 May;4(5):e115. [doi: 10.1371/journal.pmed.0040115] [Medline: 17472433]
24. Rerksuppaphol S, Rerksuppaphol L. A randomized controlled trial of zinc supplementation in the treatment of acute respiratory tract infection in Thai children. Pediatr Rep 2019 May 23;11(2):7954. [doi: 10.4081/pr.2019.7954] [Medline: 31214301]
25. Trivedi SS, Chudasama RK, Patel N. Effect of zinc supplementation in children with acute diarrhea: randomized double blind controlled trial. Gastroenterology Res 2009 Jun;2(3):168-174. [doi: 10.4021/gr2009.06.1298] [Medline: 27933128]
26. McDonald CM, Manji KP, Kisenge R, et al. Daily zinc but not multivitamin supplementation reduces diarrhea and upper respiratory infections in Tanzanian infants: a randomized, double-blind, placebo-controlled clinical trial. J Nutr 2015 Sep;145(9):2153-2160. [doi: 10.3945/jn.115.212308] [Medline: 26203094]
27. Manger MS, Taneja S, Strand TA, et al. Poor folate status predicts persistent diarrhea in 6- to 30-month-old north Indian children. J Nutr 2011 Dec;141(12):2226-2232. [doi: 10.3945/jn.111.144220] [Medline: 22013199]
28. Brown N, Roberts C. Vitamin A for acute respiratory infection in developing countries: a meta-analysis. Acta Paediatr 2004 Nov;93(11):1437-1442. [doi: 10.1080/08035250410022143] [Medline: 15513568]
29. Maggini S, Wenzlaff S, Hornig D. Essential role of vitamin C and zinc in child immunity and health. J Int Med Res 2010;38(2):386-414. [doi: 10.1177/147323001003800203] [Medline: 20515554]
30. Martineau AR, Jolliffe DA, Hooper RL, et al. Vitamin D supplementation to prevent acute respiratory tract infections: systematic review and meta-analysis of individual participant data. BMJ 2017 Feb 15;356:i6583. [doi: 10.1136/bmj.i6583] [Medline: 28202713]
31. Sazawal S, Dhingra U, Dhingra P, et al. Effects of fortified milk on morbidity in young children in north India: community based, randomised, double masked placebo controlled trial. BMJ 2007 Jan 20;334(7585):140. [doi: 10.1136/bmj.39035.482396.55] [Medline: 17132678]
32. Tam E, Keats EC, Rind F, Das JK, Bhutta AZA. Micronutrient supplementation and fortification interventions on health and development outcomes among children under-five in low- and middle-income countries: a systematic review and meta-analysis. Nutrients 2020 Jan 21;12(2):289. [doi: 10.3390/nu12020289] [Medline: 31973225]
33. Bharadva K, Mishra S, Tiwari S, et al. Prevention of micronutrient deficiencies in young children: consensus statement from Infant and Young Child Feeding Chapter of Indian Academy of Pediatrics. Indian Pediatr 2019 Jul 15;56(7):577-586. [Medline: 31333213]
34. Age structure and marital status. Census of India 2011. URL: https://censusindia.gov.in/nada/index.php/catalog/10301 [accessed 2025-06-17]
35. Sarma KVR, Udaykumar P, Balakrishna N, Vijayaraghavan K, Sivakumar B. Effect of micronutrient supplementation on health and nutritional status of schoolchildren: growth and morbidity. Nutrition 2006 Jan;22(1 Suppl):S8-14. [doi: 10.1016/j.nut.2005.07.011] [Medline: 16426962]
36. Muthayya S, Eilander A, Transler C, et al. Effect of fortification with multiple micronutrients and n-3 fatty acids on growth and cognitive performance in Indian schoolchildren: the CHAMPION (Children's Health and Mental Performance Influenced by Optimal Nutrition) Study. Am J Clin Nutr 2009 Jun;89(6):1766-1775. [doi: 10.3945/ajcn.2008.26993] [Medline: 19369376]
37. Sivakumar B, Nair KM, Sreeramulu D, et al. Effect of micronutrient supplement on health and nutritional status of schoolchildren: biochemical status. Nutrition 2006 Jan;22(1 Suppl):S15-S25. [doi: 10.1016/j.nut.2005.07.012] [Medline: 16426959]
38. Vazir S, Nagalla B, Thangiah V, Kamasamudram V, Bhattiprolu S. Effect of micronutrient supplement on health and nutritional status of schoolchildren: mental function. Nutrition 2006 Jan;22(1 Suppl):S26-S32. [doi: 10.1016/j.nut.2004.07.021] [Medline: 16426960]
39. Dietary guidelines for Indians – a manual. Indian Council of Medical Research - National Institute of Nutrition. 2011. URL: https://www.nin.res.in/downloads/DietaryGuidelinesforNINwebsite.pdf [accessed 2025-06-17]
40. Azor-Martínez E, Gonzalez-Jimenez Y, Seijas-Vazquez ML, et al. The impact of common infections on school absenteeism during an academic year. Am J Infect Control 2014 Jun;42(6):632-637. [doi: 10.1016/j.ajic.2014.02.017] [Medline: 24837113]

XSL•FO
RenderX

41. Sandora TJ, Taveras EM, Shih MC, et al. A randomized, controlled trial of a multifaceted intervention including alcohol-based hand sanitizer and hand-hygiene education to reduce illness transmission in the home. Pediatrics 2005 Sep;116(3):587-594. [doi: 10.1542/peds.2005-0199] [Medline: 16140697]
42. Barreto ML, Santos LM, Assis AM, et al. Effect of vitamin A supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in Brazil. Lancet 1994 Jul 23;344(8917):228-231. [doi: 10.1016/s0140-6736(94)92998-x] [Medline: 7913157]
43. Sur D, Gupta DN, Mondal SK, et al. Impact of zinc supplementation on diarrheal morbidity and growth pattern of low birth weight infants in Kolkata, India: a randomized, double-blind, placebo-controlled, community-based study. Pediatrics 2003 Dec;112(6 Pt 1):1327-1332. [doi: 10.1542/peds.112.6.1327] [Medline: 14654605]
44. Common Terminology Criteria for Adverse Events (CTCAE) v4.03. US Department of Health and Human Services, National Institutes of Health, National Cancer Institute. 2010. URL: https://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03/CTCAE_4.03_2010-06-14_QuickReference_8.5x11.pdf [accessed 2025-06-17]
45. Lohman TG, Roche AF, Martorell R. Anthropometric Standardization Reference Manual: Human Kinetics Books; 1988.
46. Kennedy G, Ballard T, Dop M. Guidelines for measuring household and individual dietary diversity. FAO. 2011. URL: https://www.fao.org/4/i1983e/i1983e00.pdf [accessed 2025-01-10]
47. Indian food composition tables. National Institute of Nutrition. URL: https://www.nin.res.in/ebooks/IFCT2017.pdf [accessed 2025-06-17]
48. Panda P, Sen M. Accuracy of haemoglobin estimation by non-invasive pulse co-oximetry method: a prospective observational study among neonates, children and young adults. J Med Res 2018;4(1):10-15. [doi: 10.31254/jmr.2018.4104]
49. BMI-for-age (5-19 years). World Health Organization. URL: https://www.who.int/tools/growth-reference-data-for-5to19-years/indicators/bmi-for-age [accessed 2025-06-17]
50. World Health Organization. Global Burden of Disease 2012.
51. Global burden of disease 2021. Institute for Health Metrics and Evaluation. URL: https://www.healthdata.org/sites/default/files/2024-05/GBD_2021_Booklet_FINAL_2024.05.16.pdf [accessed 2025-06-17]
52. Rosner B. Fundamentals of Biostatistics, 2nd edition: PWS Publishers; 1986:84-92.
53. Dandona R, Kumar GA, Ameer MA, Ahmed GM, Dandona L. Incidence and burden of road traffic injuries in urban India. Inj Prev 2008 Dec;14(6):354-359. [doi: 10.1136/ip.2008.019620] [Medline: 19074239]
54. Hemming K, Girling AJ, Sitch AJ, Marsh J, Lilford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. BMC Med Res Methodol 2011 Jun 30;11(102):1471-2288. [doi: 10.1186/1471-2288-11-102] [Medline: 21718530]
55. Muliyil DE, Rose A, Senthamizh SV, et al. Prevalence and risk factors of vitamin A deficiency in children and women of childbearing age in a Southern Indian tribal population: a cross-sectional study. Indian J Community Med 2019;44(2):162-165. [doi: 10.4103/ijcm.IJCM_213_18] [Medline: 31333297]
56. Awate RV, Ketkar YA, Somaiya PA. Prevalence of nutritional deficiency disorders among rural primary school children (5-15 years). J Indian Med Assoc 1997 Jul;95(7):410-411. [Medline: 9425838]
57. Gupta A, Kapil U, Ramakrishnan L, Pandey RM, Yadav CP. Prevalence of vitamin B12 and folate deficiency in school children residing at high altitude regions in India. Indian J Pediatr 2017 Apr;84(4):289-293. [doi: 10.1007/s12098-017-2291-7] [Medline: 28108882]
58. Kapil U, Sareen N. Prevalence of ferritin, folate and vitamin B12 deficiencies amongst children in 5-18 years of age in Delhi. Indian J Pediatr 2014 Mar;81(3):312. [doi: 10.1007/s12098-013-1091-y] [Medline: 23783764]
59. Sarma D, Saikia UK, Baro A. Vitamin D status of school children in and around Guwahati. Indian J Endocrinol Metab 2019;23(1):81-85. [doi: 10.4103/ijem.IJEM_552_18] [Medline: 31016159]
60. Mandlik R, Kajale N, Ekbote V, et al. Determinants of vitamin D status in Indian school-children. Indian J Endocrinol Metab 2018;22(2):244-248. [doi: 10.4103/ijem.IJEM_622_17] [Medline: 29911039]
61. Bindusha S, Riaz I, Sujith K, Kailas L. The vitamin D status in 6-14 year old children attending tertiary care teaching hospital in South India. Int J Curr Res Rev 2017;9(11):53-58. [doi: 10.7324/IJCRR.2017.91111]
62. Kapil U, Pandey RM, Goswami R, et al. Prevalence of vitamin D deficiency and associated risk factors among children residing at high altitude in Shimla district, Himachal Pradesh, India. Indian J Endocrinol Metab 2017;21(1):178-183. [doi: 10.4103/2230-8210.196031] [Medline: 28217519]
63. Kapil U, Bhadoria AS. Prevalence of folate, ferritin and cobalamin deficiencies amongst adolescent in India. J Family Med Prim Care 2014 Jul;3(3):247-249. [doi: 10.4103/2249-4863.141619] [Medline: 25374863]
64. Osei A, Houser R, Bulusu S, Joshi T, Hamer D. Nutritional status of primary schoolchildren in Garhwali Himalayan villages of India. Food Nutr Bull 2010 Jun;31(2):221-233. [doi: 10.1177/156482651003100205] [Medline: 20707228]
65. Singh MB, Fotedar R, Chalga MS, Kumar P, Deficiencies OM. Under nutrition and morbidities in school children of desert area of Rajasthan. Int J Sci Res (Ahmedabad) 2015;4(6):724-727 [FREE Full text]
66. Rao PVP, Siddhartha PV, Yeshwante P. Vitamin D deficiency, the urban epidemic: a cross-sectional study from Bhopal, India. Int J Community Med Public Health 2020;7(11):4541. [doi: 10.18203/2394-6040.ijcmph20204759]
67. Khadilkar A, Kajale N, Oza C, et al. Vitamin D status and determinants in Indian children and adolescents: a multicentre study. Sci Rep 2022 Oct 6;12(1):16790. [doi: 10.1038/s41598-022-21279-0] [Medline: 36202910]

XSL•FO
RenderX

68. Eldridge S, Ashby D, Bennett C, Wakelin M, Feder G. Internal and external validity of cluster randomised trials: systematic review of recent trials. BMJ 2008 Apr 19;336(7649):876-880. [doi: 10.1136/bmj.39517.495764.25] [Medline: 18364360]
69. Agbla SC, DiazOrdaz K. Reporting non-adherence in cluster randomised trials: a systematic review. Clin Trials 2018 Jun;15(3):294-304. [doi: 10.1177/1740774518761666] [Medline: 29608096]
70. Hahn S, Puffer S, Torgerson DJ, Watson J. Methodological bias in cluster randomised trials. BMC Med Res Methodol 2005 Mar 2;5:10. [doi: 10.1186/1471-2288-5-10] [Medline: 15743523]
71. Priest P, McKenzie JE, Audas R, Poore M, Brunton C, Reeves L. Hand sanitiser provision for reducing illness absences in primary school children: a cluster randomised trial. PLOS Med 2014 Aug;11(8):e1001700. [doi: 10.1371/journal.pmed.1001700] [Medline: 25117155]

Abbreviations

AE: adverse event
ANCOVA: analysis of covariance
CNNS: Comprehensive National Nutrition Survey
DE: design effect
FMBF: fortified malt-based food
GI: gastrointestinal
HAZ: height-for-age *z* scores
Hb: hemoglobin
ICC: intracluster correlation coefficient
IDDS: individual dietary diversity score
ITT: intention-to-treat
LAR: legally acceptable representative
mITT: modified intention-to-treat
RCT: randomized controlled trial
SPWR: study physician weekly review

Edited by Z Guo; submitted 01.11.23; peer-reviewed by A Sharma, B Shankar, N Sanwalka, SK Tchum, Y Wu; revised version received 06.04.25; accepted 07.04.25; published 25.06.25.

Please cite as:
Khadilkar A, Rawat V, Bhatt J, Chaturvedi D, Garg V, Verma P
Impact of Fortified Malt-Based Food on Immunity Outcomes in School Children in India: Cluster Randomized Controlled Trial
Interact J Med Res 2025;14:e54189
URL: https://www.i-jmr.org/2025/1/e54189
doi:10.2196/54189

© Anuradha Khadilkar, Vinay Rawat, Jaladhi Bhatt, Devyani Chaturvedi, Vivek Garg, Pankaj Verma. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 25.6.2025. 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

A Proposed Taxonomy to Holistically Classify Employee Mental Health Programs: Qualitative Taxonomy Development Study

Benedict Sevov[1,2], MSc; Robin Huettemann[1,2], Dr; Maximillian Zinner[1], Dr; Sven Meister[2,3], Prof Dr; Leonard Fehring[1,2,4], MD

[1]Faculty of Health, School of Medicine, Witten/Herdecke University, Alfred-Herrhausen-Strasse 50, Witten, Germany

[2]Health Care Informatics, Faculty of Health, School of Medicine, Witten/Herdecke University, Witten, Germany

[3]Department Healthcare, Fraunhofer Institute for Software and Systems Engineering ISST, Dortmund, Germany

[4]Department of Gastroenterology, Helios University Hospital Wuppertal, Witten/Herdecke University, Wuppertal, Germany

Corresponding Author:
Leonard Fehring, MD
Faculty of Health, School of Medicine, Witten/Herdecke University, Alfred-Herrhausen-Strasse 50, Witten, Germany

Abstract

Background: The number and diversity of employee mental health programs (EMHPs), solutions employers offer to their workforce to improve mental health, have expanded rapidly in recent years, driven by advancements in digital technology and increased global awareness of employee mental health. This dynamic has resulted in a diverse and nontransparent EMHP landscape. While existing taxonomies address specific aspects of mental health programs, a comprehensive taxonomy for classifying EMHPs in more detail remains absent. Establishing such a taxonomy would benefit researchers and practitioners by providing a common standard for categorizing EMHPs and thereby enhance transparency.

Objective: This research aimed to develop and evaluate a comprehensive taxonomy to holistically classify EMHPs, providing a practical and standardized tool for various target groups to categorize, develop, and select EMHPs.

Methods: A thorough taxonomy development process with 4 iterations was applied. The first 2 iterations used conceptual-to-empirical approaches and involved scoping reviews to identify relevant dimensions and characteristics of EMHPs. The latter 2 iterations used empirical-to-conceptual approaches and included semistructured qualitative interviews. The third iteration, involving employee interviews, aimed to identify further dimensions and characteristics of EMHPs to develop the initial taxonomy. During the fourth iteration, 17 international experts were interviewed to refine and validate the initial taxonomy. After the fourth iteration, the taxonomy was evaluated by applying it to 3 real-world EMHPs through a focus group with 5 experts to ensure that all ending conditions and the evaluation goals were met. The interrater reliability was analyzed using the proportion of observed agreement and Fleiss κ.

Results: The resulting taxonomy comprises 2 metadimensions, 21 dimensions, and 69 characteristics, offering a standardized framework for EMHP classification and analysis. Experts successfully applied the taxonomy to classify 3 selected EMHPs, resulting in an overall proportion of observed agreement of 85% and a Fleiss κ of 66%. Across dimensions, the proportion of observed agreement ranged from 64% to 100%, with Fleiss κ ranging from 20% to 100% (*P* values ranging from *P*=.004 to *P*<.001).

Conclusions: This taxonomy contributes to establishing a common standard for holistic EMHP classification. It benefits both mental health researchers and practitioners in fostering transparency and serves as a structured tool for EMHP analysis. The taxonomy enables researchers to conduct relevant future research, including the systematic identification of EMHP archetypes. In practice, the taxonomy can guide providers in market gap identification and EMHP development, inform employers in decision-making, and assist policymakers in setting up targeted support mechanisms for EMHP implementation.

(Interact J Med Res 2025;14:e67752) doi:10.2196/67752

KEYWORDS

mental health; digital mental health; workplace mental health; employee well-being; employee mental health programs; employee assistance programs; digital mental health interventions; digital mental health technologies; taxonomy; classification

XSL•FO
RenderX

Introduction

Background

Deteriorating mental health poses a major global challenge [1], particularly affecting the working population, which represents a substantial share of global society [2,3]. As a result, employee mental health has gained increasing attention from both research and society in recent years [4-6]. Employers are increasingly acknowledging the relevance of employee mental health and providing targeted support. In response, the number and diversity of employee mental health programs (EMHPs)—defined here as solutions employers offer to their employees to sustain or regain good mental health [7]—have expanded rapidly [8-11]. This evolution has been largely driven by digital technology [12-14]. Numerous new players have entered the market with technology-based EMHPs [15], including several apps, platforms, and novel artificial intelligence–enabled tools that address employees' mental well-being [16].

Transparency on available and examined programs is paramount as researchers investigate different aspects of EMHPs, such as their effectiveness and use patterns [9,17-19]. Metareviews, for example, often focus on specific EMHP types, such as digital ones [12,20,21]. Despite the presence of numerous segmentation criteria, there is still no widely accepted framework with standardized definitions to clearly describe or classify EMHPs. Furthermore, both employer representatives responsible for employee mental health and employees themselves should have access to better resources to understand what programs are available and suitable for their needs [11]. Developing a comprehensive taxonomy to categorize these programs across multiple relevant criteria would enhance transparency, support informed decision-making, and improve accessibility for all stakeholders.

Prior Work

Currently, few taxonomies exist for classifying mental health programs—particularly EMHPs—and those that do typically address only specific aspects rather than providing a comprehensive, holistic framework. To illustrate, Muñoz et al [22] present a taxonomy to classify mental health interventions by delivery mode. Similarly, Liverpool et al [23] review different digital delivery modes of youth mental health interventions. Gagnon et al [24] and Pineda et al [25] developed taxonomies with 4 and 6 dimensions, respectively, each representing a criterion or parameter, for classifying digital mental health interventions (DMHIs). These taxonomies enabled a more multidimensional approach to programs, moving beyond a single-perspective approach. Other taxonomies include Lukka and Palva's [26] classification of DMHIs based on gamification levels and the stepped care model of Ferrari et al [27] for classifying video game–based interventions targeting young people. Two additional frameworks have recently been published. Bradley et al [28] propose a classification system to guide users in choosing suitable DMHIs, using a 6-statement decision aid, in which each statement represents a relevant dimension of DMHIs. Hopkin et al [29] offer a framework with 8 dimensions for categorizing digital mental health technologies (DMHTs). Although these 2 taxonomies offer more comprehensive structures, each encompassing more than 5 dimensions, they address general digital programs and do not account for the specific needs of EMHPs. More broadly, existing taxonomies—whether focused on digital or nondigital programs—tend to overlook occupational mental health contexts and fail to consider dimensions particularly relevant for EMHPs. One exception is the work by David et al [30], which presents an approach to occupational mental health for health care workers focusing on distinct mental health stages. This framework is derived from the integrated approach to workplace mental health by LaMontagne et al [31]. However, this approach does not classify EMHPs along several dimensions. The newer taxonomy by Thomas et al [11] specifically addresses EMHPs, classifying them along 3 dimensions: primary purpose, delivery agent location, and latitude in providing. While this taxonomy provides a valuable initial structure for EMHP classification, it lacks the breadth to capture a wider range of relevant dimensions that consider other critical aspects of EMHPs.

Objectives

To address the limited comprehensiveness of existing taxonomies for EMHPs, we aimed to (1) develop a comprehensive yet concise taxonomy for classifying EMHPs along multiple dimensions, and (2) evaluate the taxonomy by applying it to 3 selected real-world EMHPs.

Methods

Study Context

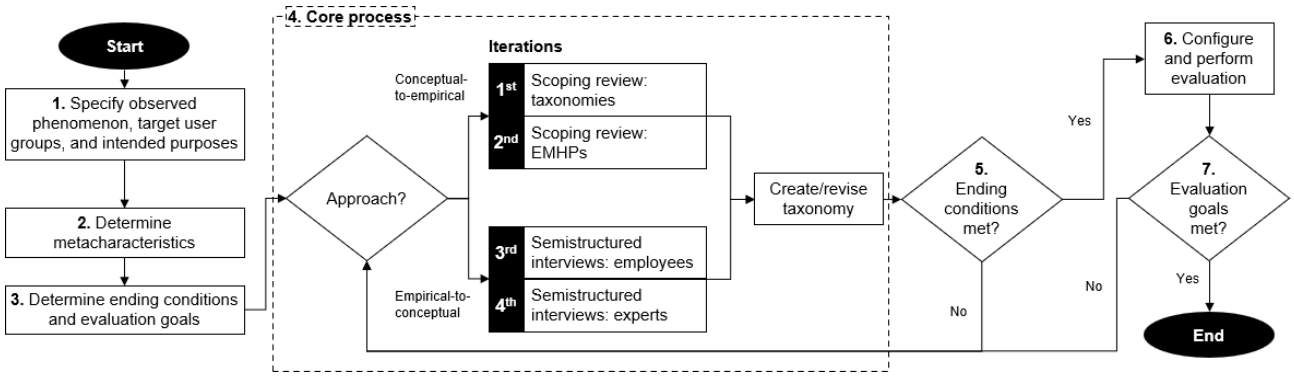
This study forms part of a broader research endeavor by the authors' research team, which investigates various facets of EMHPs, including employee preferences, use behavior, influencing factors for use, and the overall landscape and types. A clear understanding of EMHP types is essential for conducting transparent and insightful analyses of the aforementioned aspects. Given the lack of a comprehensive taxonomy that holistically classifies EMHPs, we initiated this study to develop such a framework.

Taxonomy Development Process

We followed the thorough and well-established taxonomy development process outlined by Nickerson et al [32] and the widely adopted extension proposed by Kundisch et al [33] (Figure 1). The process began with the preparation steps, including the definition of metacharacteristics, ending conditions, and evaluation goals. We then conducted the core development process, completing 4 iterations. Finally, we evaluated the resulting taxonomy.

XSL•FO
RenderX

Figure 1. Taxonomy development process according to the recommendations of Nickerson et al [32] and Kundisch et al [33]. EMHPs: employee mental health programs.



Preparation Steps

Metacharacteristics represent the foundational components of taxonomy design, including objects, dimensions, and characteristics. To ensure conceptual clarity, we defined the metacharacteristics at the outset of the study. Drawing from the research context, we defined the objects of this taxonomy as all possible EMHPs, the dimensions as the criteria or parameters by which EMHPs can be described and classified, and the characteristics as the specific attributes or options that can be attributed to each dimension. Ending conditions refer to defined states that indicate when the taxonomy development process can be concluded. We defined 5 subjective and 5 objective ending conditions based on the study by Nickerson et al [32] (Table 1).

Table . Defined subjective and objective ending conditions according to the recommendations of Nickerson et al [32].

Ending condition	Description
Subjective	
Concise	The taxonomy comprises an appropriate number of relevant dimensions to be meaningful while not overwhelming the user with an unnecessarily extensive number of dimensions.
Robust	The dimensions and characteristics of the taxonomy allow for meaningful differentiation between EMHPs[a], ensuring that these distinctions are relevant for users.
Comprehensive	All relevant EMHPs can be classified using the taxonomy and relevant characteristics of a random sample of real-world EMHPs are reflected in the taxonomy.
Extendable	New dimensions or characteristics can easily be added to the taxonomy to further evolve it if required by landscape dynamics or further research.
Explanatory	The dimensions and characteristics of the taxonomy reveal relevant and understandable information about any EMHP that is classified using the taxonomy.
Objective	
Completeness	At least 1 EMHP is classified under every characteristic of every dimension to ensure that every characteristic of the taxonomy is represented in the real world.
No addition in the last iteration	No new dimensions or characteristics were added in the last iteration (last interview of the fourth iteration); if this were not true, further validation of the last change would be required.
No merge or split in the last iteration	No dimensions or characteristics were merged or split in the last iteration (last interview of the fourth iteration); if this were not true, further validation of the last change would be required.
Uniqueness of dimensions	Every dimension is unique and not repeated, that is, no dimension duplication; otherwise, the taxonomy would contain redundancies.
Uniqueness of characteristics	Every characteristic is unique within its dimension, that is, no characteristic duplication within 1 dimension; otherwise, the taxonomy would contain redundancies.

[a]EMHP: employee mental health program.

XSL•FO
RenderX

Evaluation goals refer to specific activities a finalized taxonomy should be able to perform. We defined 3 evaluation goals in line with our research objectives and based on the study by Kundisch et al [33] (Table 2).

Table . Defined evaluation goals according to the recommendations of Kundisch et al [33].

Evaluation goal	Description
Describing	The taxonomy can be used to describe EMHPs[a] by several characteristics across multiple dimensions, meaning users can present relevant information about any EMHP using the taxonomy.
Classifying	The taxonomy can be used to classify EMHPs by assigning at least 1 characteristic per each dimension, meaning users can comprehensively describe any EMHP, presenting information for every dimension of the taxonomy.
Analyzing	The taxonomy can be used to analyze EMHPs based on similarities or differences to other EMHPs across dimensions and characteristics, meaning users can compare different EMHPs with each other and detect the most relevant similarities or differences.

[a]EMHPs: employee mental health programs.

Core Taxonomy Development Process

Overview

The first 2 iterations of the core taxonomy development process used conceptual-to-empirical approaches, involving 2 scoping reviews according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines [34] to generate first codes for the initial taxonomy. The final 2 iterations used empirical-to-conceptual approaches with semistructured qualitative interviews to identify additional codes and to refine and validate the taxonomy.

First Iteration

The first scoping review (first iteration) aimed to identify relevant taxonomies on general mental health programs, including EMHPs, and was conducted on January 2, 2024, using the following 2 search queries on PubMed:

1. ((taxonomy[MeSH Terms]) OR (framework[MeSH Terms]) OR (structure[MeSH Terms]) OR (classification[MeSH Terms]) OR (categorization[MeSH Terms])) AND ((mental health intervention[MeSH Terms]) OR (mental health program[MeSH Terms]))
2. ((taxonomy[Title/Abstract]) OR (framework[Title/Abstract]) OR (structure[Title/Abstract]) OR (classification[Title/Abstract]) OR (categorization[Title/Abstract])) AND ((mental health intervention*[Title/Abstract]) OR (mental health program*[Title/Abstract]))

The first review's PRISMA-ScR checklist [34], detailed screening process (Multimedia Appendix 1), and overview of the literature records (Multimedia Appendix 2) are provided for further details.

Second Iteration

The second scoping review (second iteration) aimed to identify relevant studies specifically on EMHPs and was initially conducted on January 12, 2023, with the literature records having been analyzed in January 2024, using the following 2 search queries on PubMed:

1. ((mental health[MeSH Terms]) OR (psychology[MeSH Terms])) AND ((employee assistance program[MeSH Terms]) OR (employer intervention[MeSH Terms]) OR (workplace intervention[MeSH Terms])) OR ((mental health[MeSH Terms]) AND (workplace[MeSH Terms]) AND (health promotion[MeSH Terms]))
2. (mental health[Title/Abstract]) AND ((employee assistance program[Title/Abstract]) OR (employer intervention[Title/Abstract]) OR (workplace intervention[Title/Abstract]))

For both reviews (first and second iterations), we followed established procedures to identify the respective search terms. We conducted initial paper searches to identify relevant papers and derive established and widely used keywords and MeSH terms. We subsequently discussed these terms within our entire author team—including SM and LF, who have extensive experience conducting scoping reviews and have published more than 20 peer-reviewed papers—prior to executing the final searches. In 2 coding rounds, authors BS and RH systematically screened the resulting papers from both reviews and assigned codes to text segments representing potential dimensions and characteristics of EMHPs. The final set of codes informed the initial taxonomy. For the second review, similar documentation is provided as for the first (Multimedia Appendices 3 and 4), detailing how codes were derived and incorporated into the taxonomy.

Third Iteration

During the third iteration, we conducted interviews with a representative sample of employees, selected to reflect diversity across characteristics, such as age, gender, and education. We recruited the participants through purposive sampling from the broader network of all authors to ensure representativeness. These interviews were conducted from February to March 2023 via phone or video calls or in person and contributed to 2 research studies conducted by our research team, while only 1 part of the factually anonymized transcripts was analyzed for this study. For this study, the purpose of the interviews was to identify relevant dimensions and characteristics of EMHPs to inform the initial taxonomy, whereas for the other study, the

XSL•FO
RenderX

purpose was to identify relevant EMHP types, as well as facilitators of and barriers to using EMHPs. The findings of the other study were published in a peer-reviewed paper on employee preference and use behavior by Sevov et al [7]. The EMHP dimensions and characteristics for this study were derived based on thematic analysis according to Braun and Clarke [35], which was conducted in 2 coding rounds by authors BS and RH. To ensure transparency and methodological rigor, we provide additional details, including the interview guide (Multimedia Appendix 5), the COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist [36], and an overview of the participant sample (Multimedia Appendix 6).

Fourth Iteration

During the fourth iteration, we conducted interviews with employee mental health experts to refine and validate the initial taxonomy. We recruited experts from 3 groups: first, relevant representatives from employers offering EMHPs (referred to as employers); second, representatives from providers of EMHPs (providers); and third, academic key opinion leaders (KOLs) who conduct research in employee mental health and related fields. To ensure geographic and cultural diversity, we systematically recruited employers and providers from multiple countries, including Germany, the United Kingdom, France, the United States, and Japan. Employers were identified through LinkedIn keyword search per country ("employee mental health" and "occupational mental health") and were contacted directly. Providers were identified through Google keyword searches in English per country ("employee mental health program provider COUNTRY"), and representatives were contacted using publicly available information. For Germany and France, these searches were additionally conducted in the respective local languages. KOLs were selected based on publication volume in employee mental health research, identified through a PubMed keyword search ("(employee mental health[Title/Abstract]) OR (employee mental health program[Title/Abstract]) OR (employee assistance program[Title/Abstract])"), and were contacted using publicly available email addresses. In total, we contacted 73 experts: 43 employers, 20 providers, and 10 KOLs.

These interviews were conducted from January to April 2024 via video calls and involved presenting and discussing the taxonomy. Using an iterative approach, we revised the taxonomy after each interview based on expert feedback and presented the updated version to the next interviewee. This process continued until both the predefined ending conditions and the minimum requirements for a representative expert sample were met, concluding after the 17th interview. The final expert sample included 5 participants from Germany, 3 from the United States, 3 from the United Kingdom, 2 from France, 2 from Japan, 1 from Sweden, and 1 from Mexico. Of these, 8 were employers, 5 were providers, and 4 were KOLs. Following the final interview, we reviewed all transcripts within the author team and made final adjustments to the taxonomy on wording based on the interview inputs to ensure clarity and consistency. Additional details related to this iteration are provided, including the interview guide (Multimedia Appendix 7), the COREQ checklist [36], and the participant sample (Multimedia Appendix 8).

Taxonomy Evaluation

Evaluation Design

After the fourth iteration, we conducted a focus group to evaluate the taxonomy—a well-established method for taxonomy evaluation [37]. Using the 3 predefined evaluation goals, 5 experts participated in the evaluation, applying the taxonomy to 3 established real-world EMHPs. The evaluation design was carefully defined based on the framework for taxonomy evaluation proposed by Szopinski et al (Figure 2) [37].

XSL•FO
RenderX

Figure 2. Applied taxonomy evaluation design based on the framework for taxonomy evaluation from the study by Szopinski et al [37].

WHO? Subject of evaluation	**Involvement in taxonomy building**	Subject has been involved in taxonomy building			Subject has not been involved in taxonomy building		
	Background	Academics			Practitioners		
	Experience	Domain			Method		
WHAT? Object of evaluation	**Type**	Real-world			About real-world		
	Involvement in taxonomy building	Object has been involved in taxonomy building		Object has been involved in taxonomy building and evolved since then		Object has not been involved in taxonomy building	
	Coverage	Exhaustive		Selective		Representative	
HOW? Method of evaluation	**Approach**	Quantitative			Qualitative		
	Method	Logical argument	Expert interview	Focus group	Survey	Delphi technique	Sorting
		Log diary	Case study	Illustrative scenario	Action research	Experiment	...

Taxonomy evaluation characteristics of this study

EMHP Selection

The selected EMHPs were established programs that had been available on the market for at least 3 years. This means that they were publicly available, for example, via a website, generally recognized, that is, among the top results in online or app store searches, and adopted and offered by multiple organizations, as evidenced by client listings on provider websites. Importantly, none of the selected EMHPs were affiliated with experts from the fourth iteration. We also selected these programs to ensure diversity: one was a fully digital online platform for employee mental health, another a purely app-based mindfulness program, and the third a program that featured digital and nondigital components focused on coaching and counseling.

Expert Selection

The experts were identified through purposive sampling to ensure that all participants were EMHP specialists with practitioner background and none of them had been involved in the fourth iteration. The selected EMHP experts represented diverse educational backgrounds, primarily in psychology or business economics, with several holding a PhD and having experience in scientific research. To ensure adequate program-specific insight, 1 representative from each of the 3 classified EMHPs participated in the focus group. During the moderated session, the experts first reflected on the taxonomy's robustness, comprehensibility, and wording. They then independently applied the taxonomy to 1 of the 3 real-world EMHPs by selecting relevant characteristics per dimension, which was followed by a group discussion on the taxonomy and the first classification [37]. After the focus group, the same experts were asked to apply the taxonomy to the remaining 2 EMHPs in written form to ensure independent and unbiased classification.

Qualitative Analysis

The qualitative reflection of the taxonomy during the focus group comprised explicit expert input on the subjective ending conditions and the understandability of the taxonomy. The experts debated whether more or less dimensions or characteristics should have been added and whether the presented characteristics were meaningful and clear to users. They weighed the advantages and disadvantages of possible changes and provided explicit and consensual conclusions on these potential adjustments. Authors BS and LF conducted the focus group and reviewed the field notes and the transcript after the session to qualitatively analyze whether the subjective ending conditions were met based on explicit expert statements and whether the experts were able to use the taxonomy to classify EMHPs based on the classification of the first EMHP and the subsequent group discussion.

Interrater Analysis

We assessed interrater reliability among the focus group experts regarding their classifications using the proportion of observed agreement and Fleiss κ [38]. We used Fleiss' approach as it represents an established extension of Cohen κ that can be applied when having a constant number of more than 2 raters assessing nominal or categorical variables, which was the case in our taxonomy evaluation [38,39]. Characteristics were coded as binary variables, that is, 1=selected, 0=not selected, to meet the mutual exclusivity condition required for Fleiss κ analysis [40]. The proportion of observed agreement and κ were calculated for each dimension. We then computed the unweighted average across dimensions to determine the overall

XSL•FO
RenderX

proportion of observed agreement and overall κ. This analysis provided empirical data that complemented the qualitative focus group discussions and supported the conclusion of the taxonomy evaluation against the predefined evaluation goals. Additional details of the evaluation process are provided, including the COREQ checklist [36] and the ACCORD (Accurate Consensus Reporting Document) checklist [41].

Ethical Considerations

The ethics committee of Witten/Herdecke University reviewed and approved this research project without raising any ethical concerns (S-317/2023). Voluntary and informed consent was obtained in advance and in writing from all participants of the qualitative interview studies and the focus group, covering both participation and publication of study findings in a peer-reviewed paper. The interview studies and the focus group adhered to established ethical scientific research standards and complied with the General Data Protection Regulation. The recordings of the qualitative interviews and the focus group session were transcribed in a factually anonymized way, ensuring that individual participants could not be identified—or only with disproportionate effort [42]. The participants of the interview studies and the focus group received no monetary compensation for their involvement.

Results

Final Taxonomy

The final taxonomy is presented as a morphological box comprising 2 metadimensions, that is, distinct categories of dimensions, 21 dimensions (D_n) and 69 characteristics ($C_{n.m}$) (Figure 3). A morphological box is defined as "a structured method for systematically looking at key characteristics or parameters of a solution [ie, dimensions] and the realistic options for each parameter [ie, characteristics]" [43]. This format offers an intuitive and systematic way to visualize the taxonomy, allowing users to mark applicable characteristics of any EMHP and generate a visual classification. To avoid misinterpretation, all dimensions are described in detail (Figure 4). The first metadimension (D_1-D_{14}) applies to all mental health programs, while the second one (D_{15}-D_{21}) captures EMHP-specific aspects. Classification across dimensions is based on the nature of the EMHP, that is, the specific program or module being classified, with several dimensions particularly relevant for differentiating digital EMHPs. Although many characteristics within these dimensions (eg, D_{10}, D_{11}, D_{13}, and D_{15}) predominantly apply to digital EMHPs, they are not exclusive to them. Following Nickerson et al [32], the taxonomy includes dimensions with mutually exclusive characteristics, meaning only 1 characteristic can apply per dimension. However, certain dimensions (D_1, D_2, D_3, D_7, D_9, and D_{16}) are not mutually exclusive, allowing multiple characteristics to be selected simultaneously to avoid loss of information and to account for the complex nature of EMHPs [44]. For example, many EMHPs address multiple mental health issues, and it may be particularly valuable to capture the specific combination of issues targeted.

Figure 3. Final taxonomy to classify employee mental health programs.

	Dimension (D_n)	Characteristics ($C_{n.m}$)						
General mental health program relevant	D_1: Purpose	$C_{1.1}$: Awareness/prevention		$C_{1.2}$: Treatment		$C_{1.3}$: Posttreatment/rehabilitation		
	D_2: Targeted mental health issue	$C_{2.1}$: Depression	$C_{2.2}$: Anxiety	$C_{2.3}$: Burnout/stress	$C_{2.4}$: Sleep disorders	$C_{2.5}$: Eating disorders	$C_{2.6}$: Substance use	$C_{2.7}$: Other
	D_3: Therapeutic approach	$C_{3.1}$: Behavioral ther. incl. cog.	$C_{3.2}$: Psychodynamic ther.	$C_{3.3}$: Pharmacological ther.	$C_{3.4}$: Other	$C_{3.5}$: None		
	D_4: Scientific evidence (M)	$C_{4.1}$: Supporting peer-reviewed studies available		$C_{4.2}$: No supporting peer-reviewed studies available				
	D_5: Focus (M)	$C_{5.1}$: Primarily mental health		$C_{5.2}$: Comprehensive well-being including mental health				
	D_6: Delivery mode (M)	$C_{6.1}$: Digital/virtual		$C_{6.2}$: Nondigital/nonvirtual		$C_{6.3}$: Hybrid		
	D_7: Interaction form	$C_{7.1}$: Self-intervention		$C_{7.2}$: Bilateral intervention		$C_{7.3}$: Group intervention		
	D_8: Human guidance/facilitation (M)	$C_{8.1}$: Unguided/self-guided		$C_{8.2}$: Partially guided		$C_{8.3}$: Fully guided		
	D_9: Qualification of human guide/facilitator	$C_{9.1}$: Officially licensed/certified	$C_{9.2}$: Nonlicensed but trained	$C_{9.3}$: Nonlicensed/nontrained	$C_{9.4}$: No human guidance/facilitation			
	D_{10}: Degree of personalization (M)	$C_{10.1}$: Personalized		$C_{10.2}$: Nonpersonalized		$C_{10.3}$: Mixed		
	D_{11}: Gamification level (M)	$C_{11.1}$: Nongamified		$C_{11.2}$: Slightly gamified		$C_{11.3}$: Strongly gamified		
	D_{12}: Activity level of user (M)	$C_{12.1}$: Interactive		$C_{12.2}$: Passive		$C_{12.3}$: Combination		
	D_{13}: Availability (M)	$C_{13.1}$: Live		$C_{13.2}$: Recorded/documented/on demand		$C_{13.3}$: Combination		
	D_{14}: Frequency (M)	$C_{14.1}$: One-time		$C_{14.2}$: Several times but limited		$C_{14.3}$: Continuous/unlimited		
EMHP specific	D_{15}: Location of use (M)	$C_{15.1}$: Location at workplace		$C_{15.2}$: Location outside of workplace		$C_{15.3}$: Independent of location		
	D_{16}: Target group level	$C_{16.1}$: Individual		$C_{16.2}$: Group		$C_{16.3}$: Organization		
	D_{17}: Targeted hierarchy level (M)	$C_{17.1}$: Employees w/o managerial responsibility		$C_{17.2}$: Managers/leaders/supervisors		$C_{17.3}$: All		
	D_{18}: Target group by profession/function (M)	$C_{18.1}$: Universal		$C_{18.2}$: Profession-specific/function-specific				
	D_{19}: Provider of solution (M)	$C_{19.1}$: Internal		$C_{19.2}$: External		$C_{19.3}$: Hybrid		
	D_{20}: Monetary cost coverage by employer (M)	$C_{20.1}$: Full coverage	$C_{20.2}$: Partial coverage	$C_{20.3}$: No coverage	$C_{20.4}$: No monetary costs involved			
	D_{21}: Recognition as work time by employer (M)	$C_{21.1}$: Fully work time	$C_{21.2}$: Partially work time	$C_{21.3}$: No work time	$C_{21.4}$: Not defined			

(M) Characteristics of dimension are mutually exclusive

XSL•FO
RenderX

Figure 4. Descriptions of all dimensions of the final employee mental health program taxonomy.

	Dimension (D_n)	Description
General mental health program relevant	D_1: Purpose	Purpose of the program, ie, which mental state of target group is addressed (eg, no vs existing vs experienced issues)
	D_2: Targeted mental health issue	Which mental health issue the program targets, ie, which concrete condition is addressed
	D_3: Therapeutic approach	How the program approaches the stated purpose, eg, sustaining or regaining good mental health through behavioral approaches
	D_4: Scientific evidence (M)	Availability of scientific evidence for the effectiveness of the program, ie, supporting peer-reviewed studies
	D_5: Focus (M)	Whether the program is primarily focused on mental health or on comprehensive well-being incl mental health alongside other aspects (eg, physical health)
	D_6: Delivery mode (M)	How the program is delivered, ie, how users access and use the program
	D_7: Interaction form	Whether and how users interact with others while using the program
	D_8: Human guidance/facilitation (M)	Whether and to what extent users are guided through the program content by a human, eg, coach, psychotherapist, or unguided
	D_9: Qualification of human guide/facilitator	Details of qualification of a human guide/facilitator, ie, level of knowledge/expertise and potential credentials
	D_{10}: Degree of personalization (M)	Whether and to what extent the program or content is tailored to individual users
	D_{11}: Gamification level (M)	Whether and to what extent gaming features are deployed in the program to address its purpose
	D_{12}: Activity level of user (M)	What activity users have to perform when using the program, eg, practicing (interactive) vs listening/watching (passive)
	D_{13}: Availability (M)	When the program content is created and/or presented or accessible, eg, live during session vs in advance and accessible anytime
	D_{14}: Frequency (M)	How often the program is delivered, ie, how often a session takes place
EMHP specific	D_{15}: Location of use (M)	Where the program can/should be used, ie, at which concrete physical location
	D_{16}: Target group level	Which level of the social ecological model the program addresses, ie, individual, group, or organization
	D_{17}: Targeted hierarchy level (M)	Which hierarchy levels or roles the program addresses, ie, individuals with or without managerial responsibility
	D_{18}: Target group by profession/function (M)	Whether the program addresses profession-specific/function-specific groups, eg, healthcare personnel, teachers
	D_{19}: Provider of solution (M)	Origin of the program, ie, where and by whom the solution is created and delivered
	D_{20}: Monetary cost coverage by employer (M)	Whether and to what extent monetary costs of the program are covered by the employer
	D_{21}: Recognition as work time by employer (M)	Whether and to what extent using the program is recognized/considered as work time by the employer

(M) Characteristics of dimension are mutually exclusive

Dimensions and Characteristics of the Taxonomy

Each iteration of the taxonomy development process contributed to the final taxonomy (Table 3). The first 3 iterations independently yielded relevant dimensions and characteristics that formed the initial taxonomy. The fourth iteration then refined the initial taxonomy and provided valuable insights. While 2 new dimensions were added, the experts consistently confirmed most dimensions and their related characteristics of the initial taxonomy. Only the wording was refined or minor changes, such as adding 1 characteristic, were made and validated during the subsequent interviews. As some dimensions sparked debates, we examined the different expert views and relied on consensus and fit with our research objectives to finally design the respective dimension.

XSL•FO
RenderX

Table . Summary of results from each of the 4 iterations.

Iteration	Key results
First iteration—scoping review: taxonomies	• 17 studies presenting relevant underlying taxonomies or frameworks • 6 taxonomies or frameworks comprising more than 2 dimensions • Only 2 taxonomies or frameworks focusing on the occupational context • 14 of the final 21 dimensions derived from this iteration • 36 of the final 69 characteristics derived from this iteration
Second iteration—scoping review: EMHPs	• 37 relevant studies on EMHPs[a] • 18 of the final 21 dimensions derived from this iteration • 48 of the final 69 characteristics derived from this iteration • 5 new dimensions derived that were not identified in the first iteration • 13 dimensions confirmed from the first iteration
Third iteration—semistructured interviews: employees	• 15 interviews with employees • 18 of the final 21 dimensions derived from this iteration • 43 of the final 69 characteristics derived from this iteration • 18 dimensions from the first and second iterations confirmed, just 1 not • No new dimension derived from this iteration
Fourth iteration—semistructured interviews: experts	• 17 interviews with experts • 2 new dimensions added during this iteration • 13 new characteristics added during this iteration, of which 6 for the 2 new dimensions • 7 dimensions renamed during this iteration • 27 characteristics renamed during this iteration

[a]EMHPs: employee mental health programs.

All 21 dimensions of the final taxonomy are presented in the following, each accompanied by a concise explanation and, where relevant, insights from the expert interviews.

The Purpose (D_1) of an EMHP defines the program's core objective. Experts debated how the characteristics reflect programs addressing chronic mental health conditions. Based on expert consensus, active management of chronic mental illness was categorized under Treatment ($C_{1.2}$), since such conditions are typically treated regardless of the temporary severity and the frequency and duration of formal treatment. The characteristics of this dimension are therefore based on the mental health stage of the individual, representing a framework that is both concise and established in the literature [10,17,45].

The Targeted mental health issue (D_2) specifies which concrete conditions an EMHP addresses. Experts discussed whether to define issues by symptoms or diagnoses. We purposely focused on diagnosable conditions according to the *International Classification of Diseases, Eleventh Revision* (*ICD-11*) by the World Health Organization while also including certain symptoms explicitly listed in *ICD-11* when particularly relevant [46]. Several experts suggested including other issues as characteristics. This represented a trade-off between robustness and conciseness, as we targeted meaningful differentiation without listing too many characteristics. While no explicit cutoff point was defined, we prioritized the characteristics presented in the final taxonomy based on the frequency of mental health issues targeted in existing EMHPs and the input from literature and the experts. Besides, geographic variations in relevance of targeted mental health issues emerged from the expert interviews: in Japan, for example, substance use was considered seldom while anger management was rated important, and in the United States, grief and loss, suicidal ideation, and neurodivergence were considered more relevant. Ultimately, these mental health issues were categorized under Other based on their lower prevalence in the analyzed EMHPs and the consolidated expert input.

The Therapeutic approach (D_3) indicates whether an EMHP is based on established therapeutic approaches, that is, clearly defined and well-researched approaches. Experts agreed on including Behavioral ($C_{3.1}$), Psychodynamic ($C_{3.2}$), and Pharmacological therapy ($C_{3.3}$), confirming cognitive behavior therapy as the most prevalent in occupational contexts, consistent with metareview findings [10,17,21,45]. As with D_2, for D_3, several other examples were stated by the experts. While single experts argued for the inclusion of holistic therapy, systemic therapy, hypnotherapy, and other concrete examples, these less prevalent approaches were categorized under Other ($C_{3.4}$). We also considered whether coaching—or related approaches such as counseling or consulting—should be included as its own therapeutic approach. We chose not to include this characteristic in the initial taxonomy, as coaching, counseling, or consulting encompasses a wide range of formats and does not represent clearly defined therapeutic approaches. Instead, coaches may draw on elements from various established therapeutic approaches in their coaching sessions, which the taxonomy can accommodate. For example, if a program offers coaching that is explicitly based on behavioral therapy, the characteristic Behavioral therapy including cognitive ($C_{3.1}$) can

XSL•FO
RenderX

be selected. In addition to the included approaches, we added the characteristic None ($C_{3.5}$) to capture EMHPs that provide information or raise awareness on employee mental health without a therapeutic approach, for example, through leaflets, books, or awareness sessions. This refinement was proposed by one expert and subsequently validated by all other experts.

The Scientific evidence (D_4) captures whether peer-reviewed studies exist for the specific EMHP, focusing on the program itself rather than its underlying therapeutic approach, since delivery can influence effectiveness. The Focus (D_5) distinguishes between EMHPs that concentrate primarily on mental health and those that address broader aspects of employee well-being, such as physical health alongside mental health. The Delivery mode (D_6) describes how users access and use the program, particularly whether digital components are involved. The Interaction form (D_7) specifies whether and how EMHP users interact with others during program use. The Human guidance/facilitation (D_8) indicates whether and to what extent EMHP use is guided by a human. The Qualification of human guide/facilitator (D_9) details the credentials of a human guide, such as officially approved licenses and certifications or professional training programs, which determine the reliability and suitability of an EMHP. The Degree of personalization (D_{10}) reflects the extent to which EMHP content or delivery is tailored to individual users. The Gamification level (D_{11}) indicates the extent to which gaming features are present in an EMHP. Although one expert raised concerns about this dimension, suggesting that gamification might trivialize mental health issues and undermine program objectives, the dimension was retained based on consensus among other experts, the growing use of gamified EMHPs [26,27], and evidence indicating that gamification can lead to higher intention to use, particularly among men [19]. The Activity level of user (D_{12}) describes the degree of activity required when using an EMHP. The Availability (D_{13}) specifies whether EMHP content is accessible on demand or bound to fixed times. The Frequency (D_{14}) captures how often an EMHP is delivered. The Location of use (D_{15}) identifies where an EMHP is used, such as the workplace or externally. Although one expert questioned this dimension's relevance, both previous research and the employee interviews suggest that the location can influence EMHP use rates due to stigmatization and anonymity concerns [47]. The Target group level (D_{16}) specifies the level of the social ecological model [48], meaning whether an EMHP targets individual employees, groups such as departments and project teams, or entire organizations. Organization-wide antistigma awareness programs are a prominent example of the latter [49]. The Targeted hierarchy level (D_{17}) describes which employees are addressed by formal role, with many EMHPs specifically targeting managers or executives, who are all subsumed under $C_{17.2}$. The Target group by profession/function (D_{18}) indicates whether an EMHP is tailored to specific professions or functions. The Provider of solution (D_{19}) identifies who develops or delivers an EMHP. The Monetary cost coverage by employer (D_{20}) specifies whether and to what extent the employer covers the monetary costs of an EMHP. The Recognition as work time by employer (D_{21}) indicates whether participation in an EMHP is counted as work time. This dimension was added based on one expert's view suggesting that this strongly influences use. For both D_{20} and D_{21}, the defining characteristics are shaped by the employer offering the EMHP rather than the program itself. Nonetheless, both dimensions were included because experts and literature highlight their relevance for use [50], and many EMHPs predefine a single characteristic by design.

Application and Evaluation of the Taxonomy

The taxonomy was evaluated through the focus group with 5 experts, which generated valuable qualitative insights from the discussion and quantitative empirical data from the interrater analysis of the classifications of the 3 established German EMHPs: the holistic digital mental health platform "Likeminded," the mindfulness mobile app "7Mind," and the "Coaching & Counseling" service from the "Fuerstenberg Institut."

At the beginning of the focus group session, the experts reflected on the taxonomy and confirmed that it provides an appropriate overview of relevant EMHP dimensions and a meaningful level of detail for analysis. The reflection further indicated that the balance between comprehensiveness and conciseness was well achieved. It also showed that users of the taxonomy need to familiarize themselves with the dimension descriptions to ensure a consistent understanding. The discussion of the first classification further demonstrated that reliable use of the taxonomy requires both a solid understanding of its dimensions and sufficient knowledge of the programs being classified.

The interrater analysis yielded a high overall proportion of observed agreement of 85%, interpreted according to Graham et al [51], and indicated a substantial strength of agreement with a κ of 66%, interpreted following Landis and Koch [52]. Across dimensions, the proportion of observed agreement ranged from 64% to 100% and κ ranged from 20% to 100%. The results were statistically significant, with P values ranging from P=.004 to P<.001.

Based on the qualitative reflections and the successful classification of the 3 EMHPs with high overall agreement measures, we concluded that the taxonomy was concise, robust, comprehensive, extendible, and explanatory, which reconfirmed that the defined subjective ending conditions were met [32]. The successful classifications also demonstrated that the taxonomy can be applied to describe, classify, and analyze EMHPs, fulfilling the predefined evaluation goals [37]. Additional details on the evaluation results, including the 3 classifications and a table with the key statistical measures of the interrater analysis per dimension, are provided (Multimedia Appendix 9).

Discussion

Principal Findings

No taxonomy is ever definitive [44] or can comprehensively capture all possible dimensions and characteristics due to the inherent complexity and evolving nature of EMHPs. Nonetheless, our proposed taxonomy provides a robust and

XSL•FO
RenderX

comprehensive framework for holistically classifying the existing and emerging landscape of EMHPs, while remaining generally extendible, which is an important aspect of a useful taxonomy [32]. To achieve this, we carefully considered trade-offs in the inclusion of dimensions and characteristics, especially between comprehensiveness, robustness, and conciseness [32]. The high overall proportion of observed agreement and the substantial interrater reliability achieved during evaluation [51,52] corroborate the robustness of our taxonomy.

Variation in the proportion of observed agreement and κ across dimensions can be partly attributed to the number of available characteristics and whether they are mutually exclusive. This is exemplified by the dimension Targeted mental health issue (D_2), which yielded the lowest proportion of observed agreement, with 64%, and the lowest κ, with 20%, and has the highest number of characteristics, that is, 7, which are not mutually exclusive. While this dimension did not lead to a fundamental discussion among focus group experts, the results indicated that experts have very subjective views on which mental health issues are addressed by a program. When multiple characteristics can be selected, classification outcomes are more likely to diverge, especially when users have varying levels of familiarity with the EMPH being assessed. Accordingly, user knowledge of an EMHP is critical for producing reliable classifications. Users should thoroughly familiarize themselves with a program before applying the taxonomy to ensure accurate and consistent results. In parallel, EMHP providers should clearly communicate the intended scope and design of their programs to support stakeholder understanding.

To ensure the comprehensiveness of our taxonomy, we included all relevant dimensions without exceeding a reasonable number. In doing so, we referred to other published taxonomies with relatively many dimensions, for example, Denecke and May [53], Scheider et al [54], and Weking et al [55] with 18, 18, and 22, respectively. We deliberately grounded our taxonomy in existing research by incorporating relevant frameworks, including the DMHI taxonomy of Gagnon et al [24], the game-based DMHI categories by Lukka and Palva [26], and the *ICD-11* by the World Health Organization [46].

Compared with other existing taxonomies, our taxonomy offers a comprehensive structure specifically designed to classify EMHPs and stands out in 2 key ways. First, with the exception of the taxonomies by Thomas et al [11] with 3 dimensions and David et al [30] and LaMontagne et al [31] with 1 dimension each, ours is to the best of our knowledge the only taxonomy explicitly developed for EMHPs rather than for general or other mental health programs. Second, while these other 3 taxonomies comprise 1-3 dimensions, our taxonomy is the only one to capture a broader and more detailed range of EMHP dimensions relevant to both academics and practitioners. This does not diminish the value and purpose of the existing taxonomies, many of which are intentionally focused on specific program scopes, dimensions, or both. The selection of an appropriate taxonomy ultimately depends on the user's objective; more targeted taxonomies may be preferable when examining a specific aspect of EMHPs or general mental health programs. However, we argue that our taxonomy encompasses most of the dimensions found in other frameworks—particularly those with only 1 or 2 dimensions—as we systematically reviewed and incorporated all relevant taxonomies and frameworks identified during the first iteration of our taxonomy development process. To demonstrate the unique value proposition of our taxonomy, we compared it with other taxonomies with more than 2 dimensions in terms of scope and value proposition (Table 4). In this comparison, we also included taxonomies published after concluding but before publishing our study. We did not compare taxonomies with 1 or 2 dimensions in more detail, as these do, by definition, not represent comprehensive taxonomies. Notably, most of the taxonomies we identified and reviewed have been published within the past 5 years.

XSL•FO
RenderX

Table . Overview of relevant taxonomies on mental health programs with more than 2 dimensions and their respective scope and value proposition.

Study	Purpose	Program scope	Focus on EMHPs[a]	Number of dimensions	Key dimensions	Value proposition
This study						
EMHP taxonomy	Providing a comprehensive taxonomy to holistically classify EMHPs	EMHPs	Yes	21	• Purpose • Therapeutic approach • Delivery mode • Interaction form • Frequency • Target group level • Provider of solution • Monetary cost coverage by employer	Comprehensive taxonomy on EMHPs covering multiple dimensions across two metadimensions: (1) General mental health program relevant, and (2) EMHP specific, thereby consolidating various dimensions from existing taxonomies
Scoping review 1: search						
Burger et al (2020) [10]	Providing an overview of the system landscape of electronic mental health interventions for depression to differentiate programs by system type	Electronic mental health interventions for depression	No	9	• Technology • Function type • Purpose • Therapy class • Quality	Overview covering dimensions describing digital mental health programs across relevant criteria and also presenting the evaluation methods used
Clay et al (2020) [56]	Providing a framework to identify and categorize core components of mental health stigma interventions in LMICs[b]	Mental health stigma reduction interventions in LMICs	No	13	• Target population • Qualifications of delivery agents • Content detail • Duration • Outcome assessments • Intervention methods • Dissemination	Comprehensive framework covering digital and nondigital mental health programs in LMICs identified in scientific studies
Gagnon et al (2022) [24]	Providing a framework to classify DMHIs[c]	DMHIs	No	4	• System • Function • Time • Facilitation	Targeted framework classifying DMHIs focusing on technology and delivery method
Lukka and Palva (2023) [26]	Providing a framework to develop game-based DMHIs	Game-based DMHIs	No	4	• Target audience • Engagement • Effectiveness	Framework supporting the development of game-based DMHIs, focusing on combining health care delivery and entertainment

XSL•FO
RenderX

Study	Purpose	Program scope	Focus on EMHPs[a]	Number of dimensions	Key dimensions	Value proposition
Pineda et al (2023) [25]	Providing an updated taxonomy to categorize literature on and different types of DMHIs for treatment and prevention	DMHIs	No	6	The taxonomy presents 6 parameters, that is, dimensions, of DMHIs, which are not labeled; the first dimension presents 4 types of DMHIs: 1. Provider administered 2. Provider administered with blended digital adjuncts 3. Self-help human supported or guided 4. Self-help fully automated	Updated taxonomy categorizing the literature on DMHIs for treatment and prevention
Scoping review 1: snowballing						
Chan et al (2015) [57]	Providing input on a possible framework to evaluate mobile mental health apps	Mobile mental health apps	No	3	All 3 dimensions with 4-5 subdimensions; • Usefulness • Usability • Integration and infrastructure	Proposed components of a framework for finding and evaluating mobile mental health apps that can be used by patients and providers
Relevant new taxonomies published after conclusion of this study						
Bradley et al (2025) [28]	Providing a classification system to support users in deciding on a suitable DMHI	DMHIs	No	6	• Target population • Function • System • Cost and access	Classification system as decision aid empowering users to make an informed choice for a suitable DMHI based on preference
Hopkin et al (2025) [29]	Providing a framework to categorize DMHTs[d]	DMHTs	No	8	• Population • Purpose • Type of approach • Human interaction • Functionality	Framework designed for various stakeholders for categorizing DMHTs and identifying key characteristics of DMHTs
Thomas et al (2024) [11]	Providing a taxonomy to organize and distinguish workplace mental health offerings	Workplace mental health offerings	Yes	3	• Purpose • Delivery agent • Latitude in providing	Taxonomy for organizing workplace mental health offerings along 3 dimensions

[a]EMHPs: employee mental health programs.

[b]LMICs: low- and middle-income countries.

[c]DMHIs: digital mental health interventions.

[d]DMHTs: digital mental health technologies.

Theoretical and Practical Implications

Our comprehensive taxonomy offers valuable contributions to both research and practice. For researchers, it provides a standardized framework for classifying EMHPs, facilitating metareviews or effectiveness assessments, as in Carolan et al [20], and for analyzing program limitations, such as the role of human guidance in digital EMHPs [47].

XSL•FO
RenderX

Practitioners also benefit from its structured approach. Providers can identify market gaps or refine their offerings based on EMHP archetypes, for example, derived from the 4 types of DMHIs proposed by Muñoz et al [22]. Developers may use the taxonomy to design more targeted and effective programs [30], while employers can apply it to guide decisions about which EMHPs to offer. Key considerations—such as whether mental health support is offered internally or through an external provider, and the associated cost [9,58]—can be examined along the taxonomy's individual dimensions. Employers can also use the taxonomy to align their EMHP selection with the characteristics of their workforce. For example, some professions are dominated by women or men, some organizations employ highly educated professionals, and others have predominantly young employees. Existing research offers insights into how sociodemographic employee groups differ in their preferences for EMHP features [7,19]. As an example, employers can use the taxonomy to design and conduct internal surveys to assess their employees' EMHP needs and preferences.

Given their potential to promote mental well-being and contribute to a mentally healthier workforce—and by extension, a healthier society—employers should be encouraged to offer EMHPs, regardless of national legislative requirements. The taxonomy is intended to support more informed and accessible decision-making for employers, while also serving as a resource for policymakers to develop targeted support mechanisms and recommendations for EMHP implementation.

Future Research

The aim of our taxonomy was to contribute to both research and the implementation of EMHPs across all types of employers, including small- and medium-sized enterprises, large corporations, and other employer types such as public organizations. Accordingly, ensuring universal applicability was a key objective, positioning the taxonomy as a potential new standard. Beyond its immediate use, the taxonomy also serves as an appropriate starting point for future research and targeted adaptations for specific application contexts.

One possibility for adaptation involves tailoring the taxonomy to specific geographic or cultural settings. Our findings indicate that the prevalence of mental health issues and the therapeutic approaches used vary across cultures. Moreover, the classification of EMHPs is influenced by cultural context. For example, the dimension Qualification of human guide/facilitator may depend on national regulatory frameworks: while some countries enforce strict certification requirements for mental health professionals, others apply more flexible standards and require lower levels of formal training [59,60]. Given the well-documented importance and effectiveness of cultural adaptation in mental health program development and implementation [61-63], we conclude that adapting taxonomies to cultural contexts is both relevant and advisable. Another possibility for adaptation is the refinement of the taxonomy to capture the role of digital technologies in EMHPs in more detail. As digital or hybrid EMHPs become increasingly prevalent [12-14], additional dimensions could be introduced to reflect the technologies used, such as AI-based systems. With the growing integration of AI-based technology in mental health care [64,65], it may be valuable to explicitly capture whether and how AI is used within EMHPs, potentially offering higher granularity.

We also see potential for the taxonomy to support future research aimed at identifying EMHP archetypes. While our evaluation revealed notable differences among the classified programs, further studies classifying a larger number of EMHPs, for example, approximately 30 objects [44], are needed to uncover consistent patterns and derive robust EMHP archetypes. Exploring the typology of digital EMHPs may be particularly relevant given the rapid emergence and evolution of new digital programs.

Limitations

Despite the thorough taxonomy development process and adherence to established research standards, this study has some typical limitations. First, the taxonomy is based on available literature and subjective expert input, requiring a balance between comprehensiveness and conciseness. As with any taxonomy, one could argue that certain dimensions or characteristics are missing or that some included elements may be less relevant. To mitigate this, we consulted a geographically diverse group of experts, ensuring broad representation and minimizing bias.

Second, although the fourth iteration included experts from 3 distinct groups and 7 countries across 3 continents, 5 of the 17 experts were from Germany. This may have introduced a geographic bias in the validation process, potentially making the taxonomy slightly more applicable for German or European EMHPs. For example, the characteristics included under Targeted mental health issue may reflect regional priorities. A taxonomy designed for other regions of the world might have included a partially different set of mental health issues, such as grief and loss, suicidal ideation, and neurodivergence in North America or anger management in parts of Asia. However, we intentionally chose not to account for country-specific systematic or institutional particularities in the development of our taxonomy. Instead, we relied on global literature and input from geographically diverse experts to determine the most relevant characteristics. We therefore remain confident in the taxonomy's universal applicability.

Third, although we conducted thorough scoping reviews, it is possible that some relevant publications were missed or that new taxonomies emerged after our research was completed, as was the case with Thomas et al [11]. To enhance overall transparency, we provide a positionality statement on our research (Multimedia Appendix 10).

Conclusions

This research contributes to establishing a new common standard for the holistic classification of EMHPs. To the best of our knowledge, the proposed taxonomy represents the first comprehensive and in-depth conceptual framework for systematically investigating the evolving landscape of EMHPs at a universal level. As such, the taxonomy offers value to both researchers and practitioners by fostering transparency and providing a shared foundation for analysis. Future research can build on this taxonomy to improve comparability across studies

XSL•FO
RenderX

and support the identification of EMHP archetypes through analyses and comparisons of classifications. In practice, program developers and employers may use the taxonomy as a decision-making tool to guide program design, selection, and implementation, while policymakers may leverage it for setting up support mechanisms for EMHP implementation.

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.

Data Availability

Data and materials of this research project, including literature record overviews and participant samples, are made available through the provided multimedia appendices. Further specific data of this research project can be shared upon reasonable request submitted to the corresponding author.

Authors' Contributions

BS and LF provided the conceptualization, design, and methodology of the study. RH, MZ, and SM contributed to the methodology of the study. BS conducted the scoping reviews; analyzed the identified literature records; planned, prepared, and conducted the qualitative interviews; transcribed and analyzed the interview transcripts; created the initial and the final taxonomy; planned, prepared, and conducted the focus group; and conducted the qualitative and the interrater analysis. RH analyzed the identified literature records and the interview transcripts. LF contributed to preparing and conducting the focus group and the qualitative and the interrater analysis. All authors reviewed the initial and the final taxonomy. BS prepared the first draft of the manuscript. All authors reviewed the manuscript. BS edited the manuscript and prepared the final version. All authors approved the final version of the manuscript. LF supervised the research project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screening process and final set of literature records of the first iteration.
[DOCX File, 107 KB - i-jmr_v14i1e67752_app1.docx]

Multimedia Appendix 2
Overview of final set of literature records of the first iteration.
[DOCX File, 69 KB - i-jmr_v14i1e67752_app2.docx]

Multimedia Appendix 3
Screening process and final set of literature records of the second iteration.
[DOCX File, 106 KB - i-jmr_v14i1e67752_app3.docx]

Multimedia Appendix 4
Overview of final set of literature records of the second iteration.
[DOCX File, 83 KB - i-jmr_v14i1e67752_app4.docx]

Multimedia Appendix 5
Relevant part of the interview guide for interviews of the third iteration.
[DOCX File, 59 KB - i-jmr_v14i1e67752_app5.docx]

Multimedia Appendix 6
Sample of interview participants (employees) of interviews of the third iteration (N=15).
[DOCX File, 61 KB - i-jmr_v14i1e67752_app6.docx]

Multimedia Appendix 7
Interview guide for interviews of the fourth iteration.
[DOCX File, 58 KB - i-jmr_v14i1e67752_app7.docx]

Multimedia Appendix 8

XSL•FO
RenderX

Sample of interview participants (experts) of interviews of the fourth iteration (N=17).
[DOCX File, 60 KB - i-jmr_v14i1e67752_app8.docx]

Multimedia Appendix 9
Results of classification of “Likeminded” platform, “7Mind” app, and “Coaching & Counseling” service of “Fuerstenberg Institut” based on the new employee mental health program taxonomy, and methods and results of the interrater analysis.
[DOCX File, 2377 KB - i-jmr_v14i1e67752_app9.docx]

Multimedia Appendix 10
Positionality statement.
[DOCX File, 57 KB - i-jmr_v14i1e67752_app10.docx]

Checklist 1
The 22-item PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist for the first iteration.
[DOCX File, 68 KB - i-jmr_v14i1e67752_app11.docx]

Checklist 2
The 22-item PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist for the second iteration.
[DOCX File, 66 KB - i-jmr_v14i1e67752_app12.docx]

Checklist 3
The 32-item COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist for interviews of the third iteration.
[DOCX File, 67 KB - i-jmr_v14i1e67752_app13.docx]

Checklist 4
The 32-item COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist for interviews of the fourth iteration.
[DOCX File, 66 KB - i-jmr_v14i1e67752_app14.docx]

Checklist 5
The 32-item COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist for the focus group.
[DOCX File, 66 KB - i-jmr_v14i1e67752_app15.docx]

Checklist 6
The 35-item ACCORD (Accurate Consensus Reporting Document) checklist for the focus group.
[DOCX File, 76 KB - i-jmr_v14i1e67752_app16.docx]

References

1. Patel V, Saxena S, Lund C, et al. Transforming mental health systems globally: principles and policy recommendations. Lancet 2023 Aug 19;402(10402):656-666. [doi: 10.1016/S0140-6736(23)00918-2] [Medline: 37597892]
2. de Oliveira C, Saka M, Bone L, Jacobs R. The role of mental health on workplace productivity: a critical review of the literature. Appl Health Econ Health Policy 2023 Mar;21(2):167-193. [doi: 10.1007/s40258-022-00761-w] [Medline: 36376610]
3. Attridge M. A global perspective on promoting workplace mental health and the role of employee assistance programs. Am J Health Promot 2019 May;33(4):622-629. [doi: 10.1177/0890117119838101c] [Medline: 31006254]
4. Strudwick J, Gayed A, Deady M, et al. Workplace mental health screening: a systematic review and meta-analysis. Occup Environ Med 2023 Aug;80(8):469-484. [doi: 10.1136/oemed-2022-108608] [Medline: 37321849]
5. Kelloway EK, Dimoff JK, Gilbert S. Mental health in the workplace. Annu Rev Organ Psychol Organ Behav 2023 Jan 23;10(1):363-387. [doi: 10.1146/annurev-orgpsych-120920-050527]
6. 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:e48504. [doi: 10.2196/48504] [Medline: 38488846]
7. Sevov B, Huettemann R, Zinner M, Meister S, Fehring L. Employee preference and use of employee mental health programs: mixed methods study. JMIR Hum Factors 2025 May 5;12:e65750. [doi: 10.2196/65750] [Medline: 40324169]

XSL•FO
RenderX

8. Jeannotte AM, Hutchinson DM, Kellerman GR. The time to change for mental health and wellbeing via virtual professional coaching: longitudinal observational study. J Med Internet Res 2021 May 13;23(7):e27774. [doi: 10.2196/27774] [Medline: 33993102]
9. Joseph B, Walker A, Fuller-Tyszkiewicz M. Evaluating the effectiveness of employee assistance programmes: a systematic review. Eur J Work Organ Psychol 2018 Jan 2;27(1):1-15. [doi: 10.1080/1359432X.2017.1374245]
10. Burger F, Neerincx MA, Brinkman WP. Technological state of the art of electronic mental health interventions for major depressive disorder: systematic literature review. J Med Internet Res 2020 Jan 20;22(1):e12599. [doi: 10.2196/12599] [Medline: 31958063]
11. Thomas BJ, Follmer KB, Meglich P. Giving organization stakeholders better help: a taxonomy for making sense of workplace mental health offerings. Group Organ Manag 2024;50(5-6):1446-1489. [doi: 10.1177/10596011241256149]
12. Armaou M, Konstantinidis S, Blake H. The effectiveness of digital interventions for psychological well-being in the workplace: a systematic review protocol. Int J Environ Res Public Health 2020;17(1):255. [doi: 10.3390/ijerph17010255]
13. Economides M, Bolton H, Male R, Cavanagh K. Feasibility and preliminary efficacy of web-based and mobile interventions for common mental health problems in working adults: multi-arm randomized pilot trial. JMIR Form Res 2022 Mar 3;6(3):e34032. [doi: 10.2196/34032] [Medline: 35238794]
14. Roche A, Kostadinov V, Cameron J, Pidd K, McEntee A, Duraisingam V. The development and characteristics of Employee Assistance Programs around the globe. J Workplace Behav Health 2018 Oct 2;33(3-4):168-186. [doi: 10.1080/15555240.2018.1539642]
15. Attridge M. The current state of employee assistance programs in the United States: a research-based commentary. Int J Sci Res Publ 2023;13(8):74-91. [doi: 10.29322/IJSRP.13.08.2023.p14010]
16. Krisher L, Boeldt DL, Sigmon CAN, Rimel SE, Newman LS. Pragmatic approach to the assessment and use of digital mental health interventions for health workers. Am J Public Health 2024 Feb;114(S2):171-179. [doi: 10.2105/AJPH.2023.307505] [Medline: 38354345]
17. Axén I, Björk Brämberg E, Vaez M, Lundin A, Bergström G. Interventions for common mental disorders in the occupational health service: a systematic review with a narrative synthesis. Int Arch Occup Environ Health 2020 Oct;93(7):823-838. [doi: 10.1007/s00420-020-01535-4] [Medline: 32246230]
18. Bouzikos S, Afsharian A, Dollard M, Brecht O. Contextualising the effectiveness of an employee assistance program intervention on psychological health: the role of corporate climate. Int J Environ Res Public Health 2022 Apr 21;19(9):5067. [doi: 10.3390/ijerph19095067] [Medline: 35564466]
19. Smail-Crevier R, Powers G, Noel C, Wang J. Health-related internet usage and design feature preference for e-mental health programs among men and women. J Med Internet Res 2019 Mar 18;21(3):e11224. [doi: 10.2196/11224] [Medline: 30882361]
20. 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]
21. Phillips EA, Gordeev VS, Schreyögg J. Effectiveness of occupational e-mental health interventions: a systematic review and meta-analysis of randomized controlled trials. Scand J Work Environ Health 2019 Nov 1;45(6):560-576. [doi: 10.5271/sjweh.3839] [Medline: 31184758]
22. Muñoz RF, Chavira DA, Himle JA, et al. Digital apothecaries: a vision for making health care interventions accessible worldwide. Mhealth 2018;4:18. [doi: 10.21037/mhealth.2018.05.04] [Medline: 30050914]
23. Liverpool S, Mota CP, Sales CMD, et al. Engaging children and young people in digital mental health interventions: systematic review of modes of delivery, facilitators, and barriers. J Med Internet Res 2020;22(6):e16317. [doi: 10.2196/16317] [Medline: 32442160]
24. Gagnon MP, Sasseville M, Leblanc A. Classification of digital mental health interventions: a rapid review and framework proposal. Stud Health Technol Inform 2022 May 25;294:629-633. [doi: 10.3233/SHTI220545] [Medline: 35612165]
25. Pineda BS, Mejia R, Qin Y, Martinez J, Delgadillo LG, Muñoz RF. Updated taxonomy of digital mental health interventions: a conceptual framework. Mhealth 2023;9(28):1-17. [doi: 10.21037/mhealth-23-6] [Medline: 37492117]
26. Lukka L, Palva JM. The development of game-based digital mental health interventions: bridging the paradigms of health care and entertainment. JMIR Serious Games 2023 Sep 4;11:e42173. [doi: 10.2196/42173] [Medline: 37665624]
27. Ferrari M, Sabetti J, McIlwaine SV, et al. Gaming my way to recovery: a systematic scoping review of digital game interventions for young people's mental health treatment and promotion. Front Digit Health 2022;4:814248. [doi: 10.3389/fdgth.2022.814248] [Medline: 35465647]
28. Bradley G, Rehackova L, Devereaux K, et al. Classifying the features of digital mental health interventions to inform the development of a patient decision aid. PLOS Digit Health 2025 Mar;4(3):e0000752. [doi: 10.1371/journal.pdig.0000752] [Medline: 40138606]
29. Hopkin G, Coole H, Edelmann F, et al. Toward a new conceptual framework for digital mental health technologies: scoping review. JMIR Ment Health 2025 Feb 19;12:e63484. [doi: 10.2196/63484] [Medline: 39969824]

XSL•FO
RenderX

30. David E, DePierro JM, Marin DB, Sharma V, Charney DS, Katz CL. COVID-19 pandemic support programs for healthcare workers and implications for occupational mental health: a narrative review. Psychiatr Q 2022 Mar;93(1):227-247. [doi: 10.1007/s11126-021-09952-5] [Medline: 34606067]
31. LaMontagne AD, Martin A, Page KM, et al. Workplace mental health: developing an integrated intervention approach. BMC Psychiatry 2014 May 9;14:131. [doi: 10.1186/1471-244X-14-131] [Medline: 24884425]
32. Nickerson RC, Varshney U, Muntermann J. A method for taxonomy development and its application in information systems. Eur J Inf Syst 2013 May;22(3):336-359. [doi: 10.1057/ejis.2012.26]
33. Kundisch D, Muntermann J, Oberländer AM, et al. An update for taxonomy designers. Bus Inf Syst Eng 2022 Aug;64(4):421-439. [doi: 10.1007/s12599-021-00723-x]
34. 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]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. 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;19(6):349-357. [doi: 10.1093/intqhc/mzm042]
37. Szopinski D, Schoormann T, Kundisch D. Because your taxonomy is worth it: towards a framework for taxonomy evaluation. 2019 Presented at: Proceedings of the European Conference on Information Systems (ECIS); Jun 8-14, 2019 URL: https://aisel.aisnet.org/ecis2019_rp/104/ [accessed 2024-12-07]
38. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull 1971;76(5):378-382. [doi: 10.1037/h0031619]
39. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Social Adm Pharm 2013;9(3):330-338. [doi: 10.1016/j.sapharm.2012.04.004] [Medline: 22695215]
40. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 1960 Apr;20(1):37-46. [doi: 10.1177/001316446002000104]
41. Gattrell WT, Logullo P, van Zuuren EJ, et al. ACCORD (Accurate Consensus Reporting Document): a reporting guideline for consensus methods in biomedicine developed via a modified Delphi. PLOS Med 2024 Jan;21(1):e1004326. [doi: 10.1371/journal.pmed.1004326] [Medline: 38261576]
42. Meyermann A, Porzelt M. Hinweise zur anonymisierung von qualitativen daten. In: Forschungsdatenzentrum (FDZ) Bildung Am DIPF (Deutsches Institut Für Internationale Pädagogische Forschung) [Guidelines for Anonymizing Qualitative Data], 1st edition 2014. URL: https://www.pedocs.de/volltexte/2022/21968/pdf/fdb-informiert_1_Meyermann_ua_Hinweise_zur_Anonymisierung_von_qualitativen_Daten_2014_v1-1_A.pdf [accessed 2024-07-12]
43. Johnson A, Gibson A. Sustainability in Engineering Design, 1st edition: Elsevier; 2014. [doi: 10.1016/B978-0-08-099369-0.00001-7]
44. Möller F, Stachon M, Azkan C, Schoormann T, Otto B. Designing business model taxonomies—synthesis and guidance from information systems research. Electron Markets 2022 Jun;32(2):701-726. [doi: 10.1007/s12525-021-00507-x]
45. Scheutzow J, Attoe C, Harwood J. Acceptability of web-based mental health interventions in the workplace: systematic review. JMIR Ment Health 2022 May 11;9(5):e34655. [doi: 10.2196/34655] [Medline: 35544305]
46. International classification of diseases 11th revision (ICD-11). World Health Organization. URL: https://icd.who.int/en [accessed 2024-07-12]
47. 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] [Medline: 29351900]
48. Wan Mohd Yunus WMA, Musiat P, Brown JSL. Systematic review of universal and targeted workplace interventions for depression. Occup Environ Med 2018 Jan;75(1):66-75. [doi: 10.1136/oemed-2017-104532] [Medline: 29074553]
49. Szeto ACH, Dobson KS. Reducing the stigma of mental disorders at work: a review of current workplace anti-stigma intervention programs. Appl Prev Psychol 2010 Jun;14(1-4):41-56. [doi: 10.1016/j.appsy.2011.11.002]
50. Dewa CS, Hoch JS. Barriers to mental health service use among workers with depression and work productivity. J Occup Environ Med 2015 Jul;57(7):726-731. [doi: 10.1097/JOM.0000000000000472] [Medline: 26147540]
51. Graham M, Milanowski A, Miller J. Measuring and promoting inter-rater agreement of teacher and principal performance ratings. 2012 URL: https://files.eric.ed.gov/fulltext/ED532068.pdf [accessed 2025-06-20]
52. 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]
53. Denecke K, May R. Developing a technical-oriented taxonomy to define archetypes of conversational agents in health care: literature review and cluster analysis. J Med Internet Res 2023 Jan 30;25:e41583. [doi: 10.2196/41583] [Medline: 36716093]
54. Scheider S, Lauf F, Geller S, Möller F, Otto B. Exploring design elements of personal data markets. Electron Markets 2023 Dec;33(1):28. [doi: 10.1007/s12525-023-00646-3]
55. Weking J, Mandalenakis M, Hein A, Hermes S, Böhm M, Krcmar H. The impact of blockchain technology on business models—a taxonomy and archetypal patterns. Electron Markets 2020 Jun;30(2):285-305. [doi: 10.1007/s12525-019-00386-3]

XSL•FO
RenderX

56. Clay J, Eaton J, Gronholm PC, Semrau M, Votruba N. Core components of mental health stigma reduction interventions in low- and middle-income countries: a systematic review. Epidemiol Psychiatr Sci 2020 Sep 4;29:e164. [doi: 10.1017/S2045796020000797] [Medline: 32883399]
57. Chan S, Torous J, Hinton L, Yellowlees P. Towards a framework for evaluating mobile mental health apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]
58. Nunes AP, Richmond MK, Pampel FC, Wood RC. The effect of employee assistance services on reductions in employee absenteeism. J Bus Psychol 2018 Dec;33(6):699-709. [doi: 10.1007/s10869-017-9518-5]
59. Baessler F, Zafar A, Gargot T, et al. Psychiatry training in 42 European countries: a comparative analysis. Eur Neuropsychopharmacol 2021 May;46:68-82. [doi: 10.1016/j.euroneuro.2021.02.003] [Medline: 33678470]
60. Asian Federation of Psychiatric Associations. The bulletin of the AFPA: the Summer 2016 Issue. Bull AFPA 2016;3(2) [FREE Full text]
61. Garabiles MR, Harper Shehadeh M, Hall BJ. Cultural adaptation of a scalable World Health Organization e-mental health program for overseas Filipino workers. JMIR Form Res 2019 Mar 29;3(1):e11600. [doi: 10.2196/11600] [Medline: 30924784]
62. Chowdhary N, Jotheeswaran AT, Nadkarni A, et al. The methods and outcomes of cultural adaptations of psychological treatments for depressive disorders: a systematic review. Psychol Med 2014 Apr;44(6):1131-1146. [doi: 10.1017/S0033291713001785] [Medline: 23866176]
63. Sit HF, Ling R, Lam AIF, Chen W, Latkin CA, Hall BJ. The cultural adaptation of step-by-step: an intervention to address depression among Chinese young adults. Front Psychiatry 2020;11:650. [doi: 10.3389/fpsyt.2020.00650] [Medline: 32733296]
64. Cross S, Bell I, Nicholas J, et al. Use of AI in mental health care: community and mental health professionals survey. JMIR Ment Health 2024 Oct 11;11:e60589. [doi: 10.2196/60589] [Medline: 39392869]
65. Olawade DB, Wada OZ, Odetayo A, David-Olawade AC, Asaolu F, Eberhardt J. Enhancing mental health with artificial intelligence: current trends and future prospects. J Med Surg Public Health 2024 Aug;3:100099. [doi: 10.1016/j.glmedi.2024.100099]

Abbreviations

ACCORD: Accurate Consensus Reporting Document
COREQ: Consolidated Criteria for Reporting Qualitative Research
DMHI: digital mental health intervention
DMHT: digital mental health technology
EMHP: employee mental health program
***ICD-11*:** *International Classification of Diseases, Eleventh Revision*
KOL: key opinion leader
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by N Cahill; submitted 20.10.24; peer-reviewed by C Curatoli, CY Chen, G Hirschfeld, I Serbanescu; accepted 31.10.25; published 18.12.25.

Please cite as:
Sevov B, Huettemann R, Zinner M, Meister S, Fehring L
A Proposed Taxonomy to Holistically Classify Employee Mental Health Programs: Qualitative Taxonomy Development Study
Interact J Med Res 2025;14:e67752
URL: https://www.i-jmr.org/2025/1/e67752
doi:10.2196/67752

© Benedict Sevov, Robin Huettemann, Maximillian Zinner, Sven Meister, Leonard Fehring. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 18.12.2025. 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

Perceptions and Experiences of Caregiver-Employees, Employers, and Health Care Professionals With Caregiver-Friendly Workplace Policy in Hong Kong: Thematic Analysis

Maggie Man-Sin Lee[1], BBA, JD, MPH, PhD; Eng-kiong Yeoh[1], MBBS; Eliza Lai-Yi Wong[1], BSN, MPH, PhD, RN; Xue Bai[2], BL, BA, PhD; Nelson Chun-Yiu Yeung[1], BSSc, MPhil, PhD; Catherine French[3], BBA, MA, PhD; Henock Taddese[4], BSc, MEd, MPH, PhD

[1]The Jockey Club School of Public Health and Primary Care, Chinese University of Hong Kong, Hong Kong, China (Hong Kong)

[2]Department of Applied Social Science, The Hong Kong Polytechnic University, Hong Kong, China (Hong Kong)

[3]Department of Health and Life Sciences, King's College London, London, United Kingdom

[4]School of Public Health, Faculty of Medicine, Imperial College London, London, United Kingdom

Corresponding Author:
Henock Taddese, BSc, MEd, MPH, PhD
School of Public Health
Faculty of Medicine, Imperial College London
St Mary's Campus
167 Medical School
London, W2 1NY
United Kingdom
Phone: 44 02075949475
Email: h.taddese@imperial.ac.uk

Abstract

Background: Caregiver-employees (CEs) for older adults experience a high burden to fulfill their dual roles. Caregiver-friendly workplace policy (CFWP) has been used in many countries to balance employment and caregiving duties, but it is a relatively new concept in Hong Kong.

Objective: This study explored the views and experiences of CEs, employers, and health care professionals regarding CFWP (specifically for older adult caregivers) in Hong Kong.

Methods: This study explored the CFWP-related views and experiences in Hong Kong using 15 in-depth interviews with purposively sampled CEs for older adults, employers, and health care professionals.

Results: Two context-related themes ("lacking leadership" and "unfavorable culture") were identified with thematic analysis. They explain the absence of CFWP in Hong Kong due to the lack of governmental and organizational leadership, and the additional burden experienced by CEs because of the working culture that underpins work-life separation, overprizing business interest, and unsympathetic corporate attitude. Implicit voice theory was applicable in explaining CEs' nondisclosure about their status at work due to potential risks. In addition, the two facilitation-related themes ("role struggle" and "inadequate support") identified in this study exhibit how the dual role had positive and negative spillover effects on each other and the inadequacy of social welfare and health care support systems.

Conclusions: We strongly recommend exploring and adopting potential CFWP in Hong Kong, considering the complexity of factors identified in this study.

(Interact J Med Res 2025;14:e58528) doi:10.2196/58528

KEYWORDS

caregiver employees; workplace; discrimination; dual roles; caregiver burden

XSL•FO
RenderX

Introduction

Background

In 2020, one billion people globally were aged 60 years or older, which is expected to double by 2050 [1]. This unprecedented increase in the aging population poses numerous economic, political, social, and health care challenges [2]. One of the pressing issues has been providing daily care to older adults. However, the size of the professional and trained workforce has not increased proportionately to meet this demand [3], driving innovations in other fields, such as robotics, to relieve some of the unmet demands of older people care [4]. Even in countries where welfare systems are well established, family caregivers provide most of the care and support for older adults [5].

Consequently, caregivers experience high burdens, strain, and poor mental health outcomes such as burnout, anxiety, and depression, as noted by several studies [6,7]. Many caregivers may simultaneously engage in paid employment, referred to as caregiver-employees (CEs), who may experience additional burdens due to high professional and caregiving demands [8]. CEs experience significant impairments [7]; they are three times more vulnerable to adverse health issues than non-CEs [9]. The progressive decline in care recipients' functional capacity requires higher physical effort and caregiving time, resulting in poor health and depression among CEs [10].

Health care professionals (HPs) have traditionally provided psycho-educational support to promote caregivers' competency and well-being [11]. In addition, employers (ERs) are increasingly adopting caregiver-friendly workplace policies (CFWPs) to mitigate some of the caregivers' burden [7]. For instance, about 80% of ERs in the United States provide some CFWP [12]. CFWP typically includes flexible working arrangements, support services, and paid or unpaid leave [13] to help CEs manage their multiple roles and improve their work-life balance [7].

The growing international commitment to sustainable organizational behavior in which employee well-being is a significant determinant of their productivity, and therefore, organizational performance has accelerated the development and widespread adoption of CFWP [14-16]. For instance, studies have shown that in addition to employee well-being, CFWP may be critical for promoting the productivity of CEs [17] as they have been associated with improved overall health of CEs by reducing occupational and overall stress, minimizing work interruptions, and improving performance [18]. There are also direct economic benefits accruing from adopting CFWP. For example, educating CEs about their caregiving activities generates a net benefit ranging from US $48,010 to US $675,657 for CEs and ERs [19].

However, the three core stakeholders of CFWP may have different perspectives on the objectives, gaps, and limitations of the existing policies against the context of, for example, personal needs (for CEs), imparting caregiving competence (for HPs), or improving organizational performance (for ERs). Thus, exploring CFWP-related experiences and perceptions of CEs' and other stakeholders, especially HPs and ERs, is imperative. While several studies have reported the types and impact of CFWP [13,17], there is a paucity of studies reporting stakeholders' perspectives on CFWPs. Although we could not identify any study exploring the perspectives of CEs, one study explored the perspectives of managers working in the Canadian health care sector [20]. In addition, we could not identify any academic literature, public policy, or articulated organizational policy specifically devised for CEs caring for older adults.

Therefore, this study explored the CFWP-related views and experiences of CEs from Hong Kong caring for older adults, along with the views and experiences of local HPs and ERs with prior experience with formulating or implementing CFWP.

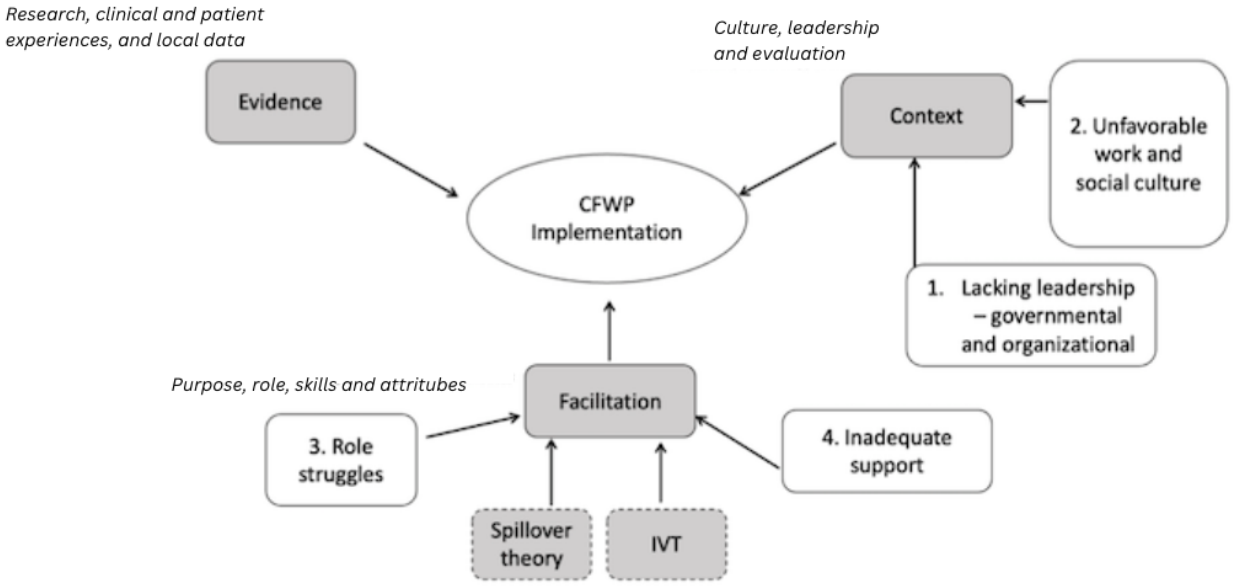
Conceptual Framework

Since CFWP is a relatively new idea in Hong Kong, we adopted the Promoting Action on Research Implementation in Health Services (PARIHS) framework for this explorative qualitative study. PARIHS was developed by the Royal College of Nursing in the United Kingdom as a conceptual framework of interacting elements to implement evidence-based practices [21]. The framework's key constructs, namely evidence, context, and facilitation, were used to guide the interview with participants regarding any future policy implementation.

To further add to the framework and explicate the facilitating factors, spillover theory [22] and implicit voice theory (IVT) [23] were used. The two theories orientate the questions from the point of view of the CEs, whereby spillover theory explores whether the interface between the microsystems of work and family is positive or negative [24], and IVT takes the gaze toward any perceived risks of inappropriateness of speaking up or disclosure of status in an organizational hierarchy [25]. These theories are fundamental in exploring the facilitation of CFWP, in other words, the expectation of and potential actions of stakeholders vis-a-vis the relatively new policy idea. Spillover theory aids the exploration of the interactions of the dual roles—work and caring—while IVT sheds light on whether CEs could or would disclose their CE status and advance their interests. Subsequently, the guiding theoretical framework adapted for this study effectively superimposes IVT and spillover theories on top of the PARIHS framework (Figure 1).

XSL•FO
RenderX

Figure 1. Schematic diagram outlining the conceptual framework of the study. CFWP: caregiver-friendly workplace policy; IVT: implicit voice theories.



Methods

Study Setting

The economic and sociocultural factors in Hong Kong put carers in a particularly precarious position. Economically, laissez-faire has long been the foundation of Hong Kong's stability and prosperity [26,27]. However, free-market neoliberalism has failed Hong Kong in terms of increasing social inequality [28,29], leaving the economically and socially vulnerable even more susceptible to discrimination and unfair treatment. In our case, the CEs care for sick older people without their dual role being recognized in the workplace. Socioculturally, the experience of CEs is complicated by filial piety, the social value of reverence, which significantly constrains the attitudes and behaviors of the caregivers since Hong Kong households emphasize moral obligation in a strictly hierarchical sense based on the recognition of aid and care given to older adults [30].

The salient economic and sociocultural context of Hong Kong is further complicated by Hong Kong's aging population, which is expected to rise sharply from 17% in 2016 to 37% by 2066 [31]. Similarly, the older people dependency ratio is expected to increase from 5:1 in 2015 to 1.8:1 by 2064 [32]. The older people dependency ratio is attributed to the high prevalence of chronic disease in the older population; about 65% of the older adults in Hong Kong have a chronic condition, while one-third have at least two chronic conditions [33]. Furthermore, Hong Kong has the fourth highest cost of living among cities globally [34]. As a result, about 59.6% of the Hong Kong population is significantly engaged in the workforce [35]. These population trends indicate a high need for CFWP in Hong Kong and underscores the urgent need to scale up CFWP in Hong Kong. This need also emerged in our recent study in which 7% of CEs in Hong Kong caring for people with Alzheimer disease had signs of probable clinical depression and 10% possible mild depression [36]. Furthermore, the responses of this study's participants indicated that CFWPs can directly improve their mental well-being and organizational performance [36].

Study Design

This study was envisioned as theory-led research, considering both conventional theories in the field of CEs study and additional selected theories fitting into the context. Spillover theory and IVT were selected to guide and inform this study's design, perspectives, and interpretative lens.

Qualitative case studies are particularly appropriate when contextual conditions believed to be relevant to the phenomenon are under investigation and when boundaries between the phenomenon and context are unclear [37]. Given the absence of CFWP in Hong Kong, the contextual conditions the CEs are subjected to are relevant to CFWP because the potential construction of CFWP is firmly based on the context at hand. Therefore, a qualitative case study methodology involving the three core stakeholders of CFWP, that is, CEs, HPs, and ERs, was used in this study. Individual in-depth semistructured face-to-face interviews were conducted to explore CEs' experiences and needs while incorporating other stakeholders' perceptions, which is well served by a naturalistic, qualitative inquiry and triangulation of perspectives [38]. The data from the interviews were subsequently coded for thematic data analysis.

Hence, the study design adopted for this research allows the issues surrounding CFWP in Hong Kong to be explored via the lenses of multiple stakeholders who may harbor diverse and potentially contrasting perspectives. This method is valuable for health science research in developing interventions (ie, CFWPs) because of its flexibility and rigor in studying complex phenomena using several data sources [37].

Participants

The study participants included three groups: CEs caring for older adults, ERs such as company management personnel, and HPs providing ancillary services for CEs. The following inclusion criteria were used for recruiting CEs: residents of Hong Kong, employed full-time, concurrently taking up a

XSL•FO
RenderX

caregiver role for either their parents or spouses aged older than 60 years, able to provide consent, and that participation would not adversely affect their health or well-being. The eligibility criteria for ER were business owners or executives of a company with at least 10 employees and at least 5 years of management experience; for HP, licensed individuals who provided health care services and had at least 5 years of experience related to caregiver service were eligible.

As the time frame for data collection was limited, a purposive sample was used to select participants who already had some CFWPs-related background knowledge or experience. Potential participants were initially identified by referrals from a scholar in a local research institute specializing in caregiver welfare, followed by a review of participants' prior participation in policy forums and public campaign platforms. This process ensured that the participants were familiarized with the concepts of CE status and the relevant rights and burdens in Hong Kong. Additional participants were identified through the snowballing technique.

Study Instrument

The semistructured in-depth interviews were conducted using a discussion guide consisting of 17 questions (Multimedia Appendix 1) related to three aspects: (1) respondents' personal experience with CEs, (2) their attitudes and preferences for policies, and (3) their perceptions of a caregiver-friendly workplace. The discussion guide was developed based on (1) a thorough literature review, (2) the master's thesis of the first author [39], and (3) the recommendation of two expert panels of researchers specializing in the field of caregivers (ELW from the Chinese University of Hong Kong and XB from the Hong Kong Polytechnic University).

The interview guide was adapted for each participant type. Thus, three sets of interview guides directed the in-depth individual interviews for the three groups of participants. The interview guide for the CEs consisted of a checklist inquiring about the basic demographics of the CEs and their care recipients, followed by open-ended questions about their personal experience on the caregiving journey, encounters on their dual roles, and their interactions with regulatory frames and other actors within the workplace context. The interview guide for ERs started with a checklist to establish basic information about the company, followed by questions about their perspective on the current situation and potential for CFWP. For HPs, the guide outlined their observations and perspectives on CE's experience and CFWP.

Interview Process

The initial interviews were conducted face-to-face in informants' offices but later switched to videoconferencing due to the COVID-19 pandemic. All interviews were audio recorded. Interviews ranged from 20 to 70 minutes. The interviews were conducted in Cantonese (Hong Kong dialect) and then translated and transcribed into English verbatim.

The participant's written consent and demographic information comprising their age, sex, working position, and marital status were obtained before the interview. During the interview, conversation starters were used to open up the participants so they would feel more at ease and provide the most candid response. Prompts were used to elicit more specific and detailed responses to fully explore their experience and perception [40,41].

Thematic Framework Analysis

Thematic framework analysis based on the conceptual guiding framework was adopted. Data coding and analysis were independently conducted by two researchers (MML and XB) to ensure the consistency and reliability of the translation from Chinese to English.

The five stages of the analytical process comprised familiarization, identification of a thematic framework, indexing, charting, and mapping and interpretation [42]. The MAXQDA 2018 analysis tool (VERBI GmbH) was used to index all transcribed data. A total of 209 indexes were charted into 4 parent themes and 65 subthemes. The last stage, mapping and interpretation, was used to gauge the prominence of key themes and subthemes across the full list of participants. While the guiding framework informed the specification of questions and the main categories or themes of interest, the themes specified and explicated by this inquiry reflect participants' views. In other words, any strategy or recommendations emerging from this research echo the participants' attitudes, beliefs, and values [43]. Data credibility was established by obtaining validation from some participants on the accuracy of how their experiences were registered in the form of interview transcripts. Some participants were also invited to review the interpretations to ensure their beliefs were accurately represented, thereby minimizing bias.

Ethical Considerations

Research ethics approval was granted by the Imperial College Research Ethics Committee and The Hong Kong Polytechnic University Department of Applied Social Science Research Committee on June 6, 2018 (reference 18IC4581) and May 1, 2018 (reference HSEARS20180413002), respectively. Written consent was obtained from interview informants after the purpose of the study, and the plan for data confidentiality, analysis, storage, and dissemination was explained to them. All data were anonymized and kept in password-protected folders accessible only to the project supervisor and the research student. Participants were not provided monetary or material compensation for participating in this study.

Results

Participant Demographics

A total of 15 participants were interviewed between May and June 2018. The CEs (n=9; CE1-9) comprised 4 female and 5 male participants aged 20-70 years from four industries (banking or finance, technology, service, and education), with caregiving experience ranging from 0.5 to 13 years (Table 1). Three of the CEs had previously quit jobs because of caregiving roles. The major conditions leading to the caregiving roles were depression, dementia, physical disabilities, and tumors.

The HPs (n=3; HP10-12) comprised a geriatrician (HP10) who has been engaged in geriatric medicine for over 35 years in

XSL•FO
RenderX

private and public practice and two social workers (HP11 and HP12) with over 5 years of experience in different contexts, such as crisis intervention, end-of-life care management, and daily support services (Table 2).

The company personnel (n=3; ER13-15) were drawn from big, medium, and small enterprises engaging in retail, law, and technology, respectively (Table 3). Only ER15 works in a company with a Family-Friendly Employers Scheme by the Hong Kong Government [44].

Table 1. Characteristics of CEs[a] included in this study.

Informant code	Sex	Age group (years)	Industry (seniority)	Condition of the care recipients	Experience as a caregiver (years)	Ever quit a job because of the caregiving role
CE1	F[b]	21-30	Bank (junior)	Brain tumor	5	No
CE2	F	51-60	Retired; bank (senior)	Depression	6	No
CE3	F	21-30	Accounting (middle)	Lung cancer	2	Yes
CE4	F	41-50	Technology	Depression	10	No
CE5	F	31-40	Primary school teacher	Joint disorder	2	No
CE6	M[c]	51-60	Security guard	Physical disabilities	0.5	No
CE7	M	Older than 60	Retired; bank (senior)	Dementia	13	Yes
CE8	M	41-50	Servicing industry (middle)	Dementia	10	No
CE9	M	41-50	Secondary school teacher	Stroke	8	Yes

[a]CE: caregiver-employees.

[b]F: female.

[c]M: male.

Table 2. Characteristics of HPs[a] included in this study.

Informant code	Profession	Years of practice
HP10	Geriatrician	>35
HP11	Social worker and gerontologist; specializes in end-of-life care	>5
HP12	Social worker; specializes in caregiver services	>5

[a]HP: health care professional.

Table 3. Characteristics of ERs[a] included in this study.

Informant code	Company size	Industry	Position	Awardee of family-friendly employers
ER13	200	Legal	Owner	No
ER14	10	Technology	Senior Manager	No
ER15	2000	Retails	Director, Human Resources	Yes

[a]ER: employer.

Current Overview

Interviews with all three CFWP stakeholder groups indicated the lack of formal workplace policy directly addressing CE issues. CEs further highlighted that any accommodation to their needs was made by management on a case-by-case, informal, and discretionary basis. Only two participants reported some form of formal workplace support for CEs. These family-friendly policies included a home office program and monthly 2-hour early leave for CE1 and CE4. These policies applied to all employees where CE1 and CE4 worked. Both CEs perceived these policies positively, exemplified by the remarks:

> *It makes a difference. It has reduced stress drastically. It is much easier to manage time, physical health (tiredness), and mental and emotional needs.* [CE4]

Thematic Analysis

Overview

The thematic analysis yielded 65 codes (Multimedia Appendix 2) categorized into four main themes: (1) lacking leadership, (2) cultural factors, (3) role struggles, and (4) inadequate support.

XSL•FO
RenderX

Lacking Leadership

A lack of guidance in laws or social policies for realizing CFWP in Hong Kong was evident during the interviews. All ER participants perceive that companies lack the resources, knowledge, and experience to support CEs. They consider government leadership essential, such as subsidies, guiding policy, and technical support. All three ER informants expressed willingness to adopt CFWP if the government takes the lead first. For example, one ER stated:

> *Government taking the lead is important. As long as there are some initiatives in the Employment Ordinance, corporates will initiate to follow...Citing paternal leave is a good example...How can he be part of the family throughout the journey? If considered, you could be surprised what the corporates would offer.* [ER15]

Within the organizational leadership, many operational factors and competing priorities influence ERs' decisions in adopting CFWP in Hong Kong, such as cost, fairness, potential uptake, and inadequate resources. ER13 reflected that caregiving roles and mental illnesses are stigmatized in Hong Kong, limiting potential service utilization. Smaller companies are hesitant to adopt CFWP officially for the potential cost implications and further responsibilities it might entail. ER14 remarked that formalization may unleash other obligations. ER14 and ER15 admitted they have other priorities, such as childcare, over CFWP. For example, ER15 highlighted, "Corporates might perceive not all older adults need help, but all children need care." HP10 confirmed that children-friendly work arrangements are common but not CFWP in Hong Kong.

Cultural Factors

CFWP aims to promote holistic integration of work and life by enabling CEs to balance their roles as employees and caregivers. However, some observations made by participants seem to suggest that the fundamental tenets of CFWP may be at odds with cultural and workplace values in the Hong Kong setting. The concept of work-life separation, rather than work-life balance, is highly prevalent in Hong Kong. A total of 6 (67%) out of 9 CEs indicated that it is an established norm in Hong Kong to separate personal or private concerns from work, if not altogether, rendering them secondary to work concerns. In other words, personal issues should not be imposed on workplaces. Failure to detach from personal stress at work could entail a reproach from management. For instance, CE7 recalled that his manager remarked, "Don't bring this burden to work. Once you clock in, you better not have this burden in your mind."

Four CE participants stated that the prevalent corporate attitude in Hong Kong prioritizes business interests over employees' well-being. The participants highlighted a culture of assumption of total work commitment from employees that leaves little room for considering CEs' concerns. They further intimated that this culture makes it difficult for CEs to raise the issue of caring for loved ones with their managers. HP10 shared this sentiment rather strongly while stating, "Employers expect slavery, working 10 hours, better not to go home, not taking leave, not getting sick, without seniors at home, no marriage, no relationship, and no children."

This attitude of management caused acute stress for employees and potential conflict with management when urgent health issues of the family care recipients need to be dealt with while the CE is at work. At these times, CEs find themselves in a dilemma of multiple roles and uncertainties. For example, one CE stated:

> *It was so stressful every time my mother called when I was at work...I was so nervous. Because of the long-term mental stress, I have hypertension.* [CE7]

All CEs lamented a general lack of compassion from their ERs and colleagues. CEs must take formal leave because "caregiving roles occur outside work hours." As a result of the unsympathetic attitude, 6 CEs chose not to disclose their CE status and remained silent, while others became more stressed or decided to quit. CE5 described her organization's management as so apathetic that she chose not to disclose her chronic condition as well. She regarded her role in the organization as "replaceable." Therefore, letting management know her caregiving role was not an option. She felt that she had no power or involvement in the decision-making process concerning her disease or her caregiving role because of the hierarchy of her organization. CE5 and HP11 attributed this passivity from the ERs to the assumption that the government should take full responsibility for supporting CEs.

CEs also spoke of colleagues' indifference to their circumstances. Fear of gossip limits CEs from disclosing their circumstances. Disclosing could be perceived as an excuse to do less work, as exemplified by a statement:

> *Trust is important, so information about me will not be spread around...I don't want people to see it as an excuse [weakness]. I don't want people to see me using my mother's condition as an excuse to do less work.* [CE3]

CE6, a security guard, believes that disclosure could dangerously affect his position in the organization. Conversely, CE4, a senior manager, expressed no hesitation or concerns regarding disclosure of his CE status. This indicates that seniority in the corporate hierarchy influences how empowered CEs may feel about disclosing their status.

The notion of the primacy of work over personal concerns is also reflected in CEs' behavior. Most CEs took it upon themselves to overcome additional burdens at work and were either not keen or could not seek help and share burdens with others. They expressed high expectations of themselves as employees and caregivers, compounding the stress. One HP noted:

> *They (CEs) will just impose the problems on themselves. They choose not to take absences but work until midnight to care for the older adults.* [HP10]

This work ethic exacerbates the stress experienced by the CEs.

Role Struggle

All CEs expressed experiencing struggles navigating the different roles as employees, caregivers, parents, partners, and friends. They felt the urge to reprioritize values in life as they

XSL•FO
RenderX

often take on a reverse role, from a child to a guardian for their frail parents. CEs reported that the caregiving role influenced their behaviors and decisions, like spending patterns or quitting a job. For instance, one CE stated:

> *I didn't want to abuse the system or the job and wanted to prioritize my students' learning experience, and that's why I decided to quit.* [CE9, teacher]

CEs sacrificed their time to undertake the caregiver role at the expense of self-care, potentially causing a loss of personal identity and difficulties connecting with themselves and others. As CE1 stressed, "The biggest change is the loss of friends and my personal time. All my time was devoted to family needs." In the case of CE7, he was abandoned by his wife after his full-time caregiving role forced him to quit his job. Thus, reprioritization of values could wreak havoc on family integrity. In addition, reprioritization may result in unmet psychological and emotional needs, introducing an identity struggle that further overwhelms the CEs. In this regard, CE7 remarked, "I need to look at myself as another person to live up to my caregiving role."

It is difficult to strike a balance between caregiver and employee roles. The intensity and stress of these roles are inseparable and nonexclusive, which may lead to a downward spiral if no support is available. New CEs may be more prone to this balancing problem because they lack institutional support and resources to prepare for these dual roles. For instance, one CE stated:

> *It was very difficult to adjust at the beginning...Letting go of work stress and switching to caregiver mode was very difficult.* [CE1]

Conversely, work could also be a protective factor for CEs to take on caregiving roles. Engagement at work can alleviate the caregivers' stress if balance can be attained. This was best exemplified by the statement:

> *If the caregiving role is 24/7, you will lose the meaning and make caregiving an obligation. It will become a burden. Especially in Hong Kong, where living space is so small. If you only take on the caregiving role, it can feel like being trapped in a cage.* [CE8]

The unavoidable familial tension—reported as "unresolved childhood emotions"—further rarifies the role struggle. HP12 spoke about how the inner child of the caregivers can be a vulnerability: "Since the care recipients are the parents who have shaped a large part of the caregivers' lives, a single gesture or action could trigger emotional stress from trauma as far [back] as those from childhood." CE5 exemplified inner child struggle as well. She reflected that she found it unfair that she had to care for her father, who had done little to support the family during her childhood. Even though she has become financially independent, CE5 would readily associate the overwhelming stress of the caregiving role with the powerlessness and lack of confidence in her child-self because of the wounds caused by her father. These emotions could intensify the present role struggle.

Inadequate Support

CEs exhibited a feeling of hopelessness mainly because of the uncertainties they faced. This hopelessness is exacerbated by the minimal assistance and support received, as explained by one HP:

> *Hopelessness...because chronic illnesses have no foreseeable end. The uncertainty of knowing how long the suffering will last causes hopelessness. The caregivers' task lists don't end either. It never ends. It is endless.* [HP11]

There is limited information to prepare and enable caregivers to live up to their roles. Besides limited information from doctors, there is a lack of reader-friendly, timely, and high-quality information. HP 10 remarked, "There is a serious need for information support...Hearsay is commonly found but useless." All CEs were frustrated in finding the quality information they needed to live up to their roles, including those related to social welfare support and handling the needs of care recipients. Most CEs had little idea about available social welfare support resources because of limited public awareness and promotion programs. In this context, one CE stated:

> *The hospital didn't actively promote them to us. And we didn't expect there is this kind of thing. It was difficult to find appropriate information. Doctors at public hospitals would not have the time to explain—only standard treatment procedures. But they didn't talk about the side effects or things about daily care to know. I was quite lost.* [CE3]

Most CEs perceived the information provided by social welfare support as inadequate. One of the social worker informants confirmed that due to potential conflicts of interest in recommending a specific service provider to caregivers, it is normal for caregivers to receive a long list of providers without recommendations, which can be confusing. The inadequate quality of information and the lack of promotional programs cause an information gap that makes CEs' caregiving journey challenging.

Accessing social welfare support is limited because of the high qualification threshold. Most cases are in the middle of the scale and ineligible for financial or physical support like respite care or home care services. In addition, the resources designated for caregivers are generally meager. One CE responded:

> *The social worker has only 1% of her work duties devoted to handling cases...she made decisions on my behalf without telling me about the rationale and the context.* [CE9]

Such negative experiences further discourage CEs from seeking social welfare support. ERs noted the need to match the services of providers with users, particular conditions they are facing, and the stage in the caring process. HP13 reported that CEs' characteristics influence service quality and satisfaction. Matching is crucial in the first stage to ensure good quality, according to ER9.

Health care providers agreed that social welfare support is inadequate because of fragmentation and lack of crisis

XSL•FO
RenderX

intervention. Crisis intervention services are available only to hospitalized cases. The chances of accessing them after discharge are minimal. The care system is fragmented and superficial because social workers have other challenging roles besides carer services. Social workers might chase the number of cases under their purview instead of going deeper individually by enhancing the caregivers' well-being or counseling them.

Besides limited resources and services, operations and priority matrix do not favor caregivers' referrals for social support, as exemplified by the following statement by HP10: "Relying exclusively on their doctors is not enough when these doctors are extremely busy." CE3 further reflected that health care providers' attitudes often lacked empathy and professionalism, resulting in disappointment and frustration. The lack of empathy or indifferent attitude of health care practitioners intensifies CEs' powerlessness. In this regard, CE3 stated, "Since time and resources are scarce in the public hospital, they wouldn't prepare you mentally for the situation. They would just tell me the symptoms and options and seek our consent for surgery immediately. There was no time allowed for our consideration. Everything was mechanical."

There are also substantial barriers to accessing professional mental health services for caregivers. For example, waiting time is considerably long, and according to HP12, "only serious cases, such as suicidal attempts, might be considered for the clinical mental health services."

Discussion

Principal Findings

This study explored the CFWP-related views and experiences of multiple stakeholders (CEs, ERs, and HPs) in Hong Kong, guided by a theoretical framework that oriented the interest toward exploring CE experiences in depth (IVT and spillover theory) and exploring the potentials of CWFP as new ideas within the broader context (PARIHS framework). Principal findings were broadly organized under two context-related ("lacking leadership" and "unfavorable culture") and two facilitation-related ("role struggle" and "inadequate support") themes.

Regarding CFWP, Hong Kong lags behind other aging Asian societies, such as Japan and Taiwan, which have already adopted 93 [45] and 21 [46] leave days for CEs, respectively. Lack of clear leadership and confusion in public policy roles and guidance was highlighted as one of the primary reasons why explicit CFWPs are absent in Hong Kong, despite the urgent need for CEs and the apparent willingness of ERs to adopt CFWP. The government taking the lead in establishing guiding frameworks for CFWP, both in law and administration, appears to be the critical missing link. Given that Hong Kong's health care system bears colonial inheritance [47], the United Kingdom's Care Act could prove instructional. The UK Care Act was implemented in 2014 to protect the rights of caregivers. The law provides provisions for the local authorities to identify CEs' needs by assessing private companies [48]. These provisions allow the UK government to strategically partner with private companies and provide support and services for ERs to achieve caregiver-friendly working environments [49]. Similar legal protection and policies adapted to the local context will be instrumental in widespread CFWP adoption among Hong Kong companies.

Regarding organizational leadership, welfare ideology and operational concerns are the two factors hindering the formulation and uptake of CFWP. Hong Kong's long-standing universal welfare ideology, described as distorted toward the government presumably providing for welfare in aging and antipoverty programs [50], fuels companies' denial of their corporate responsibility in addressing the needs of the CEs. This study also shows that operational factors, such as cost, potential uptake, and inadequate resources, prevent CFWP from gaining prominence in the organizational leadership agenda. However, the absence of CFWP is far more costly in the long run due to increased absenteeism, reduced work productivity, increased turnover, and work disruptions [51]. Sustainable long-term organizational growth would require well-considered organizational leadership that addresses this substantive issue.

In addition, local cultural factors were also identified as hindrances to CFWP adoption, such as issues concerning work-life separation, total work commitment, and lack of compassion from management. The Confucian work ethic of hard work, perseverance, and patience is also deeply embedded among Hong Kongers [52]. The sentiments and behavior reflected in our study affirm the prevalence of these values and further indicate that these may not always be compatible with the underlying principles of CFWP. The associated taboos and risks of speaking up, as perceived and experienced by CEs, further demonstrate the assertions of IVT that explain the risks and inappropriateness of speaking up.

Older adults may have diverse and unexpected health care needs that are not always associated with chronological age [53]. Thus, adopting a caregiver-friendly workplace culture conducive to compassion and acceptance of CE's circumstances is necessary. Such cultures include supportive management, a trusting environment, and establishing top-to-bottom leadership [54]. An empathic and caregiver-friendly culture would potentially encourage CEs to identify themselves (rather than implicit voice spaces), and incorporating flexibility and acceptance into business management and work culture enhances employee commitment [55].

Regarding role struggles, our findings indicate that there may be positive and negative spillovers between work and family, upholding the constructs of spillover theory. For some CEs, work can serve as a break from caregiving roles, and hence, be a protective factor, while the high burden of dual roles can force some CEs to quit work, sacrifice self-care, or even get abandoned by family members. For instance, a study focusing on the well-being of Hong Kong male caregivers reported that work has protective effects by improving caregivers' resilience and overall well-being [56]. Therefore, role-balancing is key because it rewards CEs with a sense of satisfaction and fulfillment and also avoids any ramifications by giving up paid employment for caregiving roles [57]. CFWP, which promotes and engages CEs to balance work and life, is critical for positively managing this spillover effect. These findings are

XSL•FO
RenderX

consistent with existing literature that CFWP reduces absenteeism and sick leave rates and increases employees' productivity, loyalty, engagement, and morale [58-60].

The inadequacy of support was amply reflected by CEs and HPs alike. Even though family caregiving has become more demanding, complicated, and longer-lasting than in the past, caregivers are often underprepared [61]. This study further highlights that CPWP is most needed by CEs who have just commenced the caregiving role and experience acute crisis because of little preparation and knowledge about their new role. This, in turn, emphasizes how support for new caregivers is critical for the short and long-term well-being of both CEs and their care recipients. Caregiver preparation and education are vital in reducing psychological stress and coping with the situation better [62-64]. Companies can bridge the information gap and facilitate matching according to their unique demographics and the needs of caregivers. CFWP is, therefore, imperative to help new CEs navigate their caregiving journey.

Previously, caregiving tasks in Hong Kong could be shared among siblings when family sizes were larger. However, nuclear families have become the norm in Hong Kong, with an average household size of 2.8 members [65]. Future generations are expected to have even fewer siblings sharing caregiving responsibilities, which is bound to compound and intensify the burden and sense of powerlessness. Therefore, effective CFWP is needed to ensure support for income earners to avoid the myriad adverse effects on individuals, society, and businesses.

Limitations

The major limitation of this study is that we predominantly recruited CEs engaged in full-time work, which may not fully reflect the views of CEs who are part-time or unemployed. In addition, CE participants in this study were well primed to reflect on these issues as they were already actively engaged in policy consultations for CFWP, limiting the generalizability of our findings.

Conclusions

This study is the first inquiry into the experiences of CEs and factors that may influence the adoption of CFWP in Hong Kong. The qualitative methodology and the purposively identified participants have provided a glimpse into the lives and tribulations of this otherwise invisible but growing population group. Our findings strongly indicate that Hong Kong's current workplace policy frameworks fall short of meeting the immediate needs of CEs and the long-term interests of companies and society. We urge actors to explore and adopt potential CWFP for Hong Kong, considering the complexity of factors explored in this study, including unique cultural dynamics and structural factors related to the aging population and ever-increasing health care burden. Future research should identify CEs who need the most support and pinning down the most optimal forms of regional law and organizational policy.

Acknowledgments

This manuscript did not receive any research funding.

Data Availability

The datasets generated during and/or analyzed during this study are not publicly available for privacy reasons but are available from the corresponding author upon reasonable request.

Authors' Contributions

MML, HT, XB, NCY, and CF built the study concepts and constructed the study design. MML conducted the interviews, wrote the manuscript, and provided data for all tables. MML, HT, and CF analyzed and interpreted the results. All authors reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The interview guide.
[DOCX File , 18 KB - ijmr_v14i1e58528_app1.docx]

Multimedia Appendix 2
Emergent themes, subthemes, and codes from thematic framework analysis.
[DOCX File , 14 KB - ijmr_v14i1e58528_app2.docx]

References

1. Ageing and health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2024-01-31]

XSL•FO
RenderX

2. Grinin L, Grinin A, Korotayev A. Global aging: an integral problem of the future. How to turn a problem into a development driver? In: Sadovnichy V, Akaev A, Ilyin I, Malkov S, Grinin L, Korotayev A, editors. Reconsidering the Limits to Growth. Cham: Springer International Publishing; 2023:117-135.
3. Boyle D. Caregiving within the context of elder care. In: Informal Caregivers: From Hidden Heroes to Integral Part of Care. Cham, Switzerland: Springer International Publishing; 2023:33-67.
4. Sawik B, Tobis S, Baum E, Suwalska A, Kropińska S, Stachnik K, et al. Robots for elderly care: review, multi-criteria optimization model and qualitative case study. Healthcare (Basel) 2023;11(9):1286 [FREE Full text] [doi: 10.3390/healthcare11091286] [Medline: 37174828]
5. Robison J, Shugrue N, Fortinsky RH, Gruman C. Long-term supports and services planning for the future: implications from a statewide survey of baby boomers and older adults. Gerontologist 2014;54(2):297-313. [doi: 10.1093/geront/gnt094] [Medline: 23990621]
6. Del-Pino-Casado R, Priego-Cubero E, López-Martínez C, Orgeta V. Subjective caregiver burden and anxiety in informal caregivers: a systematic review and meta-analysis. PLoS One 2021;16(3):e0247143 [FREE Full text] [doi: 10.1371/journal.pone.0247143] [Medline: 33647035]
7. Dich N, Lange T, Head J, Rod NH. Work stress, caregiving, and allostatic load: prospective results from the Whitehall II cohort study. Psychosom Med 2015;77(5):539-547 [FREE Full text] [doi: 10.1097/PSY.0000000000000191] [Medline: 25984826]
8. Bauer JM, Sousa-Poza A. Impacts of informal caregiving on caregiver employment, health, and family. Popul Ageing 2015;8(3):113-145. [doi: 10.1007/s12062-015-9116-0]
9. Yeandle S, Bennett C, Buckner L, Shipton L, Suokas A. Who Cares Wins: The Social and Business Benefits of Supporting Working Carers. London, United Kingdom: CarersUK; 2006:2016.
10. Juratovac E, Zauszniewski JA. Full-time employed and a family caregiver: a profile of women's workload, effort, and health. Womens Health Issues 2014;24(2):e187-e196. [doi: 10.1016/j.whi.2014.01.004] [Medline: 24630422]
11. Reinhard SC, Given B, Petlick NH, Bemis A. Supporting family caregivers in providing care. In: Hughes RG, editor. Patient Safety and Quality: An Evidence-Based Handbook for Nurses. Rockville, MD: Agency for Healthcare Research and Quality (US); 2008.
12. Matos K, Galinsky E, Bond JT. 2016 National study of employers. Society for human resource management. 2017. URL: https://cdn.sanity.io/files/ow8usu72/production/d73a7246cc3a3fef4ad2ece1e3d5aa4eaec2f263.pdf [accessed 2021-07-05]
13. Ireson R, Sethi B, Williams A. Availability of caregiver-friendly workplace policies (CFWPs): an international scoping review. Health Soc Care Community 2018;26(1):e1-e14 [FREE Full text] [doi: 10.1111/hsc.12347] [Medline: 27071368]
14. Vakkayil J, Torre ED, Giangreco A. “It's not how it looks!” exploring managerial perspectives on employee wellbeing. Eur Manage J 2017;35(4):548-562. [doi: 10.1016/j.emj.2016.12.002]
15. Ernst KE, Kalliath T, Kalliath T, Kalliath P. Achieving employee wellbeing in a changing work environment. Int J Manpower 2012;33(7):738-753. [doi: 10.1108/01437721211268294]
16. van de Voorde K, Paauwe J, van Veldhoven M. Employee well‐being and the HRM–organizational performance relationship: a review of quantitative studies. Int J Manage Rev 2011;14(4):391-407. [doi: 10.1111/j.1468-2370.2011.00322.x]
17. Lorenz F, Whittaker L, Tazzeo J, Williams AM. Availability of caregiver-friendly workplace policies: an international scoping review follow-up study. IJWHM 2021;14(4):459-476. [doi: 10.1108/ijwhm-10-2019-0136]
18. Ding R, Dardas A, Wang L, Williams A. Evaluation of a caregiver-friendly workplace program intervention on the health of full-time caregiver employees: a time series analysis of intervention effects. J Occup Environ Med 2020;62(10):e548-e558 [FREE Full text] [doi: 10.1097/JOM.0000000000001980] [Medline: 32769780]
19. Mofidi A, Tompa E, Williams A, Yazdani A, Lero D, Mortazavi SB. Impact of a caregiver-friendly workplace policies intervention: a prospective economic evaluation. J Occup Environ Med 2019;61(6):461-468. [doi: 10.1097/JOM.0000000000001564] [Medline: 31166280]
20. Sethi B, Williams A, Ireson R. Supporting caregiver employees: managers’ perspective in Canada. IJWHM 2017;10(1):25-41. [doi: 10.1108/ijwhm-03-2016-0016]
21. Rycroft-Malone J. The PARIHS framework—a framework for guiding the implementation of evidence-based practice. J Nurs Care Qual 2004;19(4):297-304. [doi: 10.1097/00001786-200410000-00002] [Medline: 15535533]
22. Sirgy MJ, Efraty D, Siegel P, Lee DJ. A new measure of quality of work life (QWL) based on need satisfaction and spillover theories. Social Indic Res 2001;55(3):241-302. [doi: 10.1023/a:1010986923468]
23. Knoll M, Neves P, Schyns B, Meyer B. A multi-level approach to direct and indirect relationships between organizational voice climate, team manager openness, implicit voice theories, and silence. Appl Psychol 2020;70(2):606-642. [doi: 10.1111/apps.12242]
24. Hill E, Ferris M, Märtinson V. Does it matter where you work? A comparison of how three work venues (traditional office, virtual office, and home office) influence aspects of work and personal/family life. J Vocational Behav 2003;63(2):220-241. [doi: 10.1016/s0001-8791(03)00042-3]
25. Milliken FJ, Morrison EW, Hewlin PF. An exploratory study of employee silence: issues that employees don’t communicate upward and why*. J Manage Stud 2003;40(6):1453-1476. [doi: 10.1111/1467-6486.00387]

XSL•FO
RenderX

26. Chan MK. Stability and prosperity in Hong Kong: the twilight of laissez-faire colonialism?—A review article. J Asian Stud 2011;42(3):589-598. [doi: 10.2307/2055519]
27. Chiu SWK, Siu KYK. Hong Kong as an economic miracle? The myth of laissez-faire and industrialization. In: Hong Kong Society. Singapore: Springer Nature Singapore; 2022:137-163.
28. Tang G, Yuen RHY. Hong Kong as the 'Neoliberal Exception' of China: transformation of Hong Kong citizenship before and after the transfer of sovereignty. J Chin Polit Sci 2016;21(4):469-484. [doi: 10.1007/s11366-016-9438-7]
29. Peck J. Milton's paradise: situating Hong Kong in neoliberal lore. JLPE 2021;1(2):189-192. [doi: 10.5070/lp61251592]
30. Ng WI, Che SL, Li X, Zhu M. Association of filial attitude, filial behavior and death literacy: implications for development of death system in Guangdong-Hong Kong-Macao Greater Bay area of China. BMC Public Health 2024;24(1):721 [FREE Full text] [doi: 10.1186/s12889-024-18197-3] [Medline: 38448863]
31. Hong Kong population projections. Census and Statistics Department (Hong Kong Special Administrative Region Government). 2017. URL: https://www.censtatd.gov.hk/en/data/stat_report/product/FA100061/att/B71710FA2017XXXXB0100.pdf [accessed 2021-07-05]
32. Population policy. Chief Secretary for Administration's Office (Hong Kong Special Administrative Region Government). 2015. URL: https://gia.info.gov.hk/general/201501/15/P201501150880_0880_140608.pdf [accessed 2021-07-05]
33. Wong SYS, Zou D, Chung RY, Sit RW, Zhang D, Chan D, et al. Regular source of care for the elderly: a cross-national comparative study of Hong Kong With 11 developed countries. J Am Med Dir Assoc 2017;18(9):807.e1-807.e8. [doi: 10.1016/j.jamda.2017.05.009] [Medline: 28647578]
34. Worldwide cost of living 2022: how soaring inflation has affected prices globally. The Economist Intelligence Unit Limited. 2022. URL: https://www.eiu.com/n/wp-content/uploads/2022/11/Worldwide-Cost-of-Living-2022.pdf [accessed 2024-02-01]
35. Labour force 2021. Census and Statistics Department (Hong Kong Special Administrative Region Government). URL: https://www.censtatd.gov.hk/en/web_table.html?id=7 [accessed 2021-07-05]
36. Lee MMS, Yeoh EK, Wang K, Wong ELY. Caregiver employees' mental well-being in Hong Kong. Healthcare (Basel) 2024;12(10):1013 [FREE Full text] [doi: 10.3390/healthcare12101013] [Medline: 38786422]
37. Yin RK. Case Study Research Design and Methods. Thousand Oaks, CA: Sage; 2014:282.
38. Flick U. Doing Triangulation and Mixed Methods. London, United Kingdom: SAGE Publications Ltd; 2018.
39. Lee M, French C, Bai X, Yeung C, Taddese H. A Case Study of Caregiver-Friendly Workplace Policy in Hong Kong [master's thesis]. London, UK: Imperial College London; Jun 2018:1-20.
40. Service RW. Techniques and procedures for developing grounded theory. In: Corbin J, Strauss A, editors. Organizational Research Methods. Thousand Oaks, CA: SAGE Publications, Inc; 2008:614-617.
41. Crewell JW, Crewell JD. Research Design—Qualitative, Quantitative, and Mixed Methods Approaches. Washington, DC: Sage Publication; 2017.
42. Ritchie J, Spencer L. Qualitative data analysis for applied policy research. In: Byrman AG, Burgess R, editors. Analyzing Qualitative Data. London, United Kingdom: Routledge; 1994:173-194.
43. Srivastava A, Thomson SB. Framework analysis: a qualitative methodology for applied policy research. Journal of Administration and Governance 2009;4:72-79 [FREE Full text]
44. Family-friendly employers award scheme 2017/18. Family Council. 2022. URL: https://www.familycouncil.gov.hk/en/Employment/employers2017.html [accessed 2022-03-14]
45. Nishimoto M. Choices of leave when caring for family members: What is the best system for balancing family care with employment. Jpn Labor Rev 2017;14(1):45-67 [FREE Full text]
46. Family caregivers. Taiwan Association of Family Caregivers. 2022. URL: https://www.familycare.org.tw/policy/901 [accessed 2022-04-18]
47. Kong X, Yang Y, Gao J, Guan J, Liu Y, Wang R, et al. Overview of the health care system in Hong Kong and its referential significance to mainland China. J Chin Med Assoc 2015;78(10):569-573 [FREE Full text] [doi: 10.1016/j.jcma.2015.02.006] [Medline: 26144022]
48. Ahmed R. Care act FAQ. Carers UK. 2022. URL: https://www.carersuk.org/help-and-advice/guides-and-tools/your-guide-to-the-care-act-england/ [accessed 2022-03-10]
49. Hockin H. Partner with us. Carers UK. 2022. URL: https://www.carersuk.org/how-you-can-help/partner-with-us [accessed 2022-03-10]
50. Yu A. The failure of the welfare ideology and system in Hong Kong: a historical perspective. Hum Aff 2021;31(1):99-108. [doi: 10.1515/humaff-2021-0009]
51. Keating NC, Fast JE, Lero DS, Lucas SJ, Eales J. A taxonomy of the economic costs of family care to adults. J Econ Ageing 2014;3:11-20. [doi: 10.1016/j.jeoa.2014.03.002]
52. Wong D. Stress factors and mental health of carers with relatives suffering from schizophrenia in Hong Kong: implications for culturally sensitive practices. Br J Social Work 2000;30(3):365-382. [doi: 10.1093/bjsw/30.3.365]
53. World report on ageing and health. World Health Organization. 2015. URL: https://apps.who.int/iris/handle/10665/186463 [accessed 2024-02-01]
54. Bernard M, Phillips JE. Working carers of older adults. Commun Work Fam 2007;10(2):139-160. [doi: 10.1080/13668800701270075]

XSL•FO
RenderX

55. Dex S. Families and work in the twenty-first century. 2003. URL: https://www.researchgate.net/publication/239987709_Families_and_work_in_the_twenty-first_century [accessed 2024-02-01]
56. Siu OL. Research on family-friendly employment policies and practices in Hong Kong: implications for work-family interface. In: Lu L, Cooper C, editors. Handbook of Research on Work-Life Balance in Asia. Cheltenham, United Kingdom: Edward Elgar Publishing; 2015.
57. Horrell B, Breheny M, Stephens C. Capability to care and work: when dual roles intersect. Vulnerable Groups Inclusion 2014;5(1):24357. [doi: 10.3402/vgi.v5.24357]
58. Barroso M. EuroWORK case study: work adjustments, Axa, Portugal. Eurofound. 2011. URL: https://www.eurofound.europa.eu/en/case-studies-0 [accessed 2024-02-01]
59. Peters P. EuroWORK case study: awareness raising, Novel Zorg and Wonen, the Netherlands. Eurofound. 2011. URL: https://www.eurofound.europa.eu/en/case-studies-0 [accessed 2024-02-01]
60. Humer Z. EuroWORK case study: awareness raising, Mercator, Slovenia. Eurofound. 2011. URL: https://www.eurofound.europa.eu/en/case-studies-0 [accessed 2024-02-01]
61. Committee on Family Caregiving for Older Adults, Board on Health Care Services, Health and Medicine Division, National Academies of Sciences, Engineering, and Medicine. In: Schulz R, Eden J, editors. Families Caring for an Aging America. Washington, DC: National Academies Press (US); 2016.
62. Schumacher KL, Stewart BJ, Archbold PG, Caparro M, Mutale F, Agrawal S. Effects of are giving demand, mutuality, and preparedness on family caregiver outcomes during cancer treatment. Oncol Nurs Forum 2008;35(1):49-56. [doi: 10.1188/08.ONF.49-56] [Medline: 18192152]
63. Williamson GM, Schulz R. Caring for a family member with cancer: past communal behavior and affective reactions. J Appl Soc Pyschol 2006;25(2):93-116. [doi: 10.1111/j.1559-1816.1995.tb01586.x]
64. Wang YN, Shyu YIL, Tsai WC, Yang PS, Yao G. Exploring conflict between caregiving and work for caregivers of elders with dementia: a cross-sectional, correlational study. J Adv Nurs 2013;69(5):1051-1062. [doi: 10.1111/j.1365-2648.2012.06092.x] [Medline: 22776026]
65. Population and household statistics. Census and Statistics Department (Hong Kong Special Administrative Region Government). 2021. URL: https://gia.info.gov.hk/general/202103/30/P2021033000302_363913_1_1617079632759.pdf [accessed 2024-02-01]

Abbreviations

CE: caregiver-employee
CFWP: caregiver-friendly workplace policy
ER: employer
HP: health care professional
IVT: implicit voice theory
PARIHS: Promoting Action on Research Implementation in Health Services

Edited by T de Azevedo Cardoso; submitted 18.03.24; peer-reviewed by D Huang; comments to author 17.10.24; revised version received 03.11.24; accepted 19.12.24; published 10.02.25.

Please cite as:
Lee MMS, Yeoh EK, Wong ELY, Bai X, Yeung NCY, French C, Taddese H
Perceptions and Experiences of Caregiver-Employees, Employers, and Health Care Professionals With Caregiver-Friendly Workplace Policy in Hong Kong: Thematic Analysis
Interact J Med Res 2025;14:e58528
URL: https://www.i-jmr.org/2025/1/e58528
doi:10.2196/58528
PMID:

©Maggie Man-Sin Lee, Eng-kiong Yeoh, Eliza Lai-Yi Wong, Xue Bai, Nelson Chun-Yiu Yeung, Catherine French, Henock Taddese. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 10.02.2025. 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

Understanding the Mental and Physical Burdens of Physicians and Identifying Support Interventions in Bangladesh: Qualitative Study

Rahat Jahangir Rony[1], PhD; Shams Akbar Aalok[2], BSc; Lamia Amin Tisha[2], BSc; Marzan Mahatab[2], BSc; Nova Ahmed[2], PhD

[1]Department of Computer Science and Informatics, School of Computer Science, Cardiff University, Cardiff, United Kingdom

[2]Department of Electrical and Computer Science, North South University, Dhaka, Bangladesh

Corresponding Author:
Rahat Jahangir Rony, PhD
Department of Computer Science and Informatics
School of Computer Science
Cardiff University
Senghennydd Road
Cathays
Cardiff, CF24 4AG
United Kingdom
Phone: 44 07311249074
Email: ronyr@cardiff.ac.uk

Abstract

Background: The COVID-19 pandemic had a substantial, negative impact on the world, and physicians played a crucial role in providing health care while facing the risk of contracting the virus themselves. While working on the frontlines, they also needed to protect themselves and their families from the virus. Unfortunately, their mental health was not given the attention it deserved. Many physicians experienced burnout due to the numerous challenges they faced, yet they received little support. Resource-limited countries such as Bangladesh were particularly affected due to a lack of resources. Although high-income countries have proposed a well-being model for physicians, this model is not directly applicable to resource-limited nations. However, redefining the model to suit the specific needs of physicians in resource-limited countries could provide sustainable support for their well-being.

Objective: We aimed to gain a deeper understanding of the mental and physical burdens faced by Bangladeshi physicians during the COVID-19 pandemic, and the contextual factors influencing their well-being. By understanding these aspects, we can recommend an adaptable, effective, and sustainable contextual model.

Methods: We conducted semistructured online interviews with 14 physicians in Chattogram, Bangladesh, during the COVID-19 pandemic. The physicians actively working in the COVID-19 unit were recruited from public and private hospitals through purposive sampling. Participants were aged between 25 and 35 years and had up to 8 years of working experience, including 43% (6/14) interns, 36% (5/14) medical officers, 14% (2/14) researchers, and 7% (1/14) surgeons. Each interview was conducted in Bengali, and we obtained consent to record the audio. Overall, 637 minutes of discussion were translated and transcribed. The results were analyzed using reflexive thematic analysis.

Results: We identified factors that impacted physicians' mental and physical health and well-being during the COVID-19 pandemic. They frequently dealt with undiagnosed patients, which put them at risk. Physicians often feared the potential danger their profession posed to their families, choosing to prioritize their family's safety over their own. In addition, heavy workloads, excessive duty hours, and a shortage of colleagues substantially affected their sleep patterns and disrupted their regular work schedules. Instead of receiving societal support, they often faced negative perceptions from the public. In addition, during times of mass patient deaths, many physicians struggled to cope with their emotions without any mental health support.

Conclusions: Our work shows physicians' mental and physical health burdens with various contextual difficulties. We understood these concerns and suggested a contextual (emphasizes understanding and addressing users' behavior within its specific context) intervention model inspired by the well-being framework. We emphasize the importance of integrating both contextual and

XSL•FO
RenderX

technological interventions. Through this model, our goal is to involve stakeholders in redesigning the work environment for physicians, ensuring it is sustainable in the long term and adaptable to different situations.

(Interact J Med Res 2025;14:e76934) doi:10.2196/76934

KEYWORDS

COVID-19 pandemic; Bangladeshi doctors; doctors' mental burden; doctors' physical burden; doctors' well-being; doctors' workload; doctors' families health; well-being model; contextual support and social awareness; technology interventions; artificial intelligence; AI

Introduction

Background

The COVID-19 pandemic has led to a significant loss of human life globally [1-5], with physicians serving as frontline warriors dedicated to the well-being of the population while trying to keep themselves and their families safe [6,7]. Tragically, many physicians lost their lives while performing their professional duties. In the United States, over 4500 physicians succumbed to SARS-CoV-2 [8], while approximately 6500 health care professionals died in England and Wales in the United Kingdom [9]. These tragic figures were not isolated to high-income nations. The crisis extended to India, with approximately 1800 deaths of health care workers [10], and China reported the loss of 64 health care workers [11], among many other countries. The situation was similarly dire in Bangladesh, a resource-limited country, which reported around 113 physicians' deaths within just a few months in 2020 [12]. With about 80,000 physicians caring [12] for a population of 180 million in Bangladesh, the high patient-to-physician ratio highlighted the immense pressure physicians face. This pandemic placed additional strain on physicians and the health care system, resulting in significant psychological distress [13].

While numerous studies have discussed the health effects of COVID-19 on patients, the impact on physicians has received less attention. Physician burnout became a prominent worldwide issue during the pandemic [14-16]. This burnout stemmed from several factors, including stress [17,18], workload [19,20], anxiety and depression [18,21], inadequate support [17,18,20], lack of protective gear [13,20], and many more. Research indicated that 32.9% of physicians experienced stress, 34.9% dealt with depression, and 39.5% faced anxiety [22]. In addition, a study by Emre et al [23] found that among 225 physicians, 17.8% exhibited symptoms of posttraumatic stress disorder, regardless of their gender, marital status, and other factors.

Factors contributing to this distress included a high risk of infection, inadequate protection from exposure, excessive workload, frustration, discrimination, isolation, negative emotions from patients, lack of family contact, and overall exhaustion [24,25]. Physicians often had to maintain social distance from their families to reduce the risk of transmitting the virus, further impacting their mental health [13,26]. In addition, during the pandemic, the extreme workload due to a surge in patients resulted in increased stress and depression [20,21]. Covering for colleagues [19,27] is a duty hour complexity [19] that also contributed to the increased workload. Because of these high work demands (mental, emotional, physical, and temporal), limited recovery experiences (such as relaxation and detachment from work) impact physicians' mental health [18]. Many physicians also experienced sleep disturbances, with more than two-thirds reporting insomnia (68.3%) and a majority (93.7%) experiencing stress as a result [15]. Despite risking their lives serving patients, physicians received inadequate support from hospitals, government agencies, and society at large [20,28]. A lack of logistics support, insufficient backing from peers and supervisors, and low feelings of occupational competence during COVID-19–related tasks contributed to the emotional toll on physicians [18].

Much of the existing research has focused on physicians' experiences in high-income countries, overlooking the burdens and well-being of physicians in resource-limited nations. Although some issues are similar, conditions are often worse in resource-limited countries, which struggle significantly with resource shortages and inadequate administrative support [20,29,30]. The pandemic has exacerbated the existing shortcomings within these health care systems, creating additional challenges for physicians. Therefore, there is an urgent need to explore the mental and physical health of physicians in the context of resource-limited countries and the unique problems they face.

Aims and Approach

This research aimed to understand the mental and physical burdens faced by Bangladeshi physicians during the COVID-19 pandemic, considering the specific contextual challenges they encountered. We sought to explore how they delivered care to patients while also managing their personal and professional lives, the sociocontextual problems they faced during the pandemic, and how these issues impacted their overall well-being. However, there is limited research on how physicians can sustain their well-being. Daniels et al [31] presented a well-being model for health care professionals during the pandemic, which identified several key factors necessary for maintaining well-being, including access to basic physical resources, communication, embedded support systems, and psychological interventions. This model inspired us, and a similar approach tailored to address the specific concerns and contextual problems faced by physicians in Bangladesh would enhance its effectiveness and sustainability. In this paper, we report the results of in-depth interviews conducted with physicians in Bangladesh.

Methods

We conducted in-depth qualitative interviews with 14 physicians in Bangladesh during the COVID-19 pandemic to understand

XSL•FO
RenderX

their well-being and stress indicators. The interviews were semistructured and held online.

Research Settings and Self-Disclosure

All participants were located in Chattogram, Bangladesh, which is the country's largest port city and second-major city, situated approximately 300 km southeast of the capital, Dhaka. During the pandemic, Chattogram experienced a high frequency of COVID-19 infections, prompting us to recruit physicians from this area.

It is important for the readers to understand the researchers' perspectives [32]. We chose this location because we are familiar with it. All researchers (2 men and 3 women) were born and raised in Bangladesh, making them well-acquainted with the local cultural norms. As native speakers, the researchers have conducted multidisciplinary research related to health and technology. These experiences allowed us to effectively facilitate this study.

Recruitment and Participants

We focused on recruiting physicians who worked in both public and private hospitals during the COVID-19 pandemic. We used a purposive sampling method [33] to recruit 14 participants to accomplish this. During recruitment, we used close connections, such as relatives, friends, and colleagues, to identify potential participants. We explained our inclusion criteria (physicians who worked as frontliners in hospitals during the intense period of the pandemic) to our contacts, who helped us connect with eligible individuals. Then, we contacted them by phone to explain our research objectives and interview techniques. Subsequently, we formally recruited interested participants.

The 14 recruited participants were aged between 25 and 35 years. They had between 1 and 8 years of work experience in public and private hospitals in Bangladesh. In total, 12 (86%) participants were unmarried, while the remaining 2 (14%) were married. The group included 6 (43%) intern physicians (who had recently completed their professional degrees), 5 (36%) medical officers (serving as junior physicians), 2 (14%) researchers (focused on medical topics), and 1 (7%) surgeon (a senior physician). A detailed description of the participants has been discussed in the Results section.

Although the sample size was small, we reached a saturation level within that sample because we conducted in-depth interviews. After each interview, we organized and analyzed the data to improve our approach for the next one. Regardless of the participants' gender, job positions, workplace, or years of experience, we found that they shared similar experiences during their duties throughout the pandemic, confirming the saturation of data.

Ethical Considerations

This research was approved by the ethics committee at North South University, Bangladesh (2021/OR-NSU/IRB/0303). Before the interviews, we provided participants with soft copies of consent forms that included the research objectives, aims, and methods in Bengali. We clearly explained that the interviews would be audio recorded and that we would collect demographic information, such as name, age, gender, designation, and workplace. We assured participants that all information would be anonymized to protect their identities. They returned the consent forms online to confirm their participation before the interviews began.

We avoided using participants' names, home addresses, and institutional affiliations as identifiers to maintain confidentiality. Instead, participants were assigned code names, such as participant 1, participant 2, ..., and participant 14. They were also informed that they had the complete right to withdraw from the interview at any time without losing their incentives. Only the researchers had access to the data, which was securely stored on a locked drive. These data will not be shared or transferred to any third-party entities. After the interviews, all participants received notebooks and stationery items sent via courier as a token of appreciation for their participation.

Procedure and Study Design

We conducted interviews between May and December 2021. The participants had busy schedules due to their duty hours; in addition, to maintain social distancing due to the emergence of COVID-19 during that time, we scheduled the interviews online based on preferences and availability. Three researchers were present at each interview to collect data, as the interviews were semistructured. We used the Google Meet platform for the interviews, and most lasted approximately 60 (mean 45, SD 16.26) minutes or more. Each of the 12 participants was interviewed individually, with only participants 5 and 6 attending together as acquaintances. The interviews were conducted in Bengali, audio recorded with the participant's consent, and we followed the same protocols for both individual and group interviews.

The interviews began with an introduction and a discussion about the participants' general work experiences during the first wave (March 2020-May 2020) and second wave (March 2021-May 2021) of the COVID-19 pandemic, as this was a new situation for them. We then explored their mental health and the factors that contributed to their stress during both normal and pandemic periods. In addition, we were interested in understanding how the pandemic impacted their social, family, and work lives, as well as what kind of support they received or felt was lacking, which could enhance their overall well-being. Finally, we inquired about any potential technology that could support their well-being. The participants were responsive and willingly shared their perceptions and suggestions.

Analysis

The sessions were conducted in Bengali and required transcription and translation of all the interviews. Being fluent in both English and Bengali, all authors transcribed and translated the audio carefully directly to English (forward translation). A total of 637 minutes of audio recordings resulted in a rich corpus of narrative data comprising 156 pages (approximately 71,700 words). All authors revised the translated version of the transcription that assured the validity of the data and analyzed it all without using generative artificial intelligence (AI). We created high-level summaries of narrative data for the following interviews. After completing all the interviews, we

XSL•FO
RenderX

used reflexive thematic analysis, both inductive and deductive [34,35]. While we had specific topics of interest, such as physicians' stress, work schedules, sleeping patterns, social problems, and lack of support, we used open inductive coding to identify broader trends [36,37]. We read and reread all the transcriptions multiple times to become thoroughly familiar with the data and subsequently identified 52 different codes across all the transcriptions. We then organized the codes using MIRO, mapping them out, and creating groups of similar codes. We kept in mind the specific topics, and lastly, during theme generation, we adopted a deductive approach that was guided by the model by Daniels et al [31], allowing us to structure our findings within an established theoretical framework. In total, 3 themes were generated.

Theme 1 focused on the risks and fears that physicians and their families faced regarding the SARS-CoV-2 virus as they worked on the front lines with patients who had undiagnosed infections, which ultimately led to their own infections. *Theme 2* highlighted the increased stress and workload that physicians experienced, indicating that the COVID-19 pandemic negatively impacted their overall well-being. In addition, the social isolation they faced contributed to their mental decline. *Theme 3* addressed the emotional connections that physicians formed with their patients while on duty during the pandemic, as well as the lack of mental support to help them cope with the emotional toll and potential mental breakdowns.

Results

Participants Demography

Table 1 shows participants demography, where among the 14 participants, 12 (86%) were female and 2 (14%) were male. We recruited more females for 2 reasons. One, the number of registered female physicians is more (52%) than the male physicians (48%) [38]. Second, women in Bangladesh in general face greater challenges in the workplace due to patriarchal social norms in Bangladesh, which was also the case during the COVID-19 pandemic [38,39]. However, we did not observe any gender-based differences regarding their work experiences during the COVID-19 pandemic (Table 1).

Table 1. Participants details (N=14).

Participant code	Marital status	Designation	Format	Unit and type of hospital
First round of individual interviews				
Participant 1	Single	Intern physicians	Videoconferencing	Medicine unit, public hospital
Participant 2	Single	Medical officer	Videoconferencing	Public hospital
Participant 3	Married	Medical officer	Videoconferencing	Influenza corner, public hospital
Participant 4	Single	Medical officer	Videoconferencing	Post–COVID-19 unit, public hospital
Only group interview				
Participant 5	Married	Medical officer	Videoconferencing	Cardiology unit, public hospital
Participant 6	Single	Doctor, involved in medical research in their own hospitals	Videoconferencing	University medicine department, public university
Second round of individual interviews				
Participant 7	Single	Medical officer	Videoconferencing	Medicine unit, public hospital
Participant 8	Single	Doctor, involved in medical research in their own hospitals	Videoconferencing	Public hospital
Participant 9	Single	Intern doctor	Videoconferencing	Public hospital
Participant 10	Single	Assistant surgeon	Videoconferencing	Private medical hospital
Participant 11	Single	Intern doctor	Videoconferencing	Medicine unit, public hospital
Participant 12	Single	Intern doctor	Videoconferencing	Medicine unit, public hospital
Participant 13	Single	Intern doctor	Videoconferencing	Patriarchal surgery unit, public hospital
Participant 14	Single	Intern doctor	Videoconferencing	Cardiology unit, private hospital

Theme 1: COVID-19 Pandemic's Domino Effect: Undiagnosed Cases Impact Physicians and Their Families Health

The Risk From Undiagnosed COVID-19 Cases

During the peak period of the COVID-19 pandemic, the participant physicians primarily attended to undiagnosed patients in emergency zones of hospitals, often unaware of whether these patients were COVID-19 positive, putting the physicians at risk. The situation consisted of cases where patients were unaware of the conditions, and in a few cases, it was deliberate denial.

Many patients did not follow safety precautions while visiting hospitals and were not cautious about their own health conditions. This was due to the high physician-to-patient ratio along with a coping mechanism of handling fear, as a few

XSL•FO
RenderX

participants shared. The physicians diagnosed these patients by listening to their concerns and assessing their symptoms, often influenza-like and other respiratory disease symptoms. On the basis of their evaluations, they performed rapid antigen tests to obtain results in about 15 minutes, after which they transferred patients who were suspected to have COVID-19 to the COVID-19 unit for further care. Participant 10 expressed the following:

> *Most of the time, patients would come to us undiagnosed. They didn't even know their bodies were carrying the germs of COVID-19 and were unaware of it.*

All the physicians noted that patients often concealed their symptoms. This puts the physicians in a difficult position regarding the decision-making of whether to send patients to the COVID-19 pandemic unit. Once in the COVID-19 pandemic unit, patients were generally suspected of having the virus, and due to the lack of personal safety measures, there was a risk of further spread within the hospital. Consequently, physicians suggested that patients return home. As participant 13 explained the following,

> *We tell them that if they stay in the hospital during the COVID-19 period, they might get infected.*

Conversely, when physicians suspected COVID-19 pandemic symptoms—such as breathing difficulties, overlapping infections, or Omicron variant symptoms such as diarrhea and drops in oxygen levels—patients frequently denied they had COVID-19 and only accepted the diagnosis when test results came back positive. This situation increased the risk of physicians getting infected during the assessment period. Participant 6 expressed frustration about the patient's behavior:

> *People didn't admit they had COVID-19. They would say they had stomachaches and wouldn't mention a sore throat. After examining them, we noticed they had a fever and were coughing. After testing, we found them to be COVID-19 positive. Many of our doctors were affected by these patients.*

Some patients had clear signs of COVID-19 but yielded negative test results. As a precaution, physicians often tested themselves 2 to 3 times due to the potential risk of infection. Many physicians reported being infected with COVID-19, yet they continued to work for the well-being of their patients, often sacrificing their own health. Participant 13 shared her experience:

> *One night, I received seven patients, all suspected to have COVID-19. I also felt sick and distressed, but I didn't pay much attention and did my duty. It was nearly impossible to maintain proper sterilization and disinfection. I could feel something was happening to me, but I ignored it.*

It showed the major difficulty physicians faced, constantly battling the fear inside them.

Between Fear and Safety: Families Navigating the COVID-19 Pandemic Threat

The participant physicians often stayed away from families and did not share their concerns around getting infected with the SARS-CoV-2 virus. Several participants were infected with COVID-19 but chose to ignore their conditions due to concern for their families. If they suspected they had contracted the virus, they often refrained from sharing this information, especially with their parents, fearing it would cause unnecessary panic, as participant 1 expressed:

> *I can't share these feelings with my family, like telling them I suspect I could be affected. They panic easily. Indeed, people can't handle panic like a doctor can.*

Their primary concern was the fear of transmitting the virus to family members after returning home from work. Physicians who had previously contracted COVID-19 or whose loved ones had experienced it were particularly cautious while on duty, which heightened their tension and fear. Participant 9 noted the following:

> *Our colleagues who had COVID-19 or whose families suffered heavily due to it took extensive precautions when they returned to the ward. They struggled to breathe, wore N95 masks layered with double surgical masks, and worried about not wanting to be on duty for too long out of fear.*

These physicians felt an overwhelming burden regarding the risk their profession posed to their families, often prioritizing their families' safety over their own, as they shared during the interviews. When family members fell ill, they felt an added responsibility to secure the best medical care for them, which only increased their stress. Some physicians intentionally maintained isolation to minimize interaction with their families to keep them safe. Despite their families expressing concern about their health, these physicians tried alleviating fears. As participant 12 explained the following:

> *My family used to stress about it. But I wasn't as fearful because I didn't expose them to what I was going through. I felt like a carrier, so I lived in the hostel.*

This way, the physicians managed to safeguard their loved ones while managing their own health concerns.

Physicians' Struggles of Fighting COVID-19 Along With Families

During the COVID-19 pandemic, the families of physicians also faced significant challenges, as many were infected with the virus. Physicians struggled to balance their hospital duties with their responsibilities at home despite ensuring safety. For instance, participant 3 shared her experience of her baby, aged 14 months, contracting COVID-19 due to her exposure. At the same time, her husband was hospitalized for breathing issues, illustrating the helplessness, stress, and fear she felt regarding her family's situation, especially as she could not maintain proper isolation with a young child:

> *Despite maintaining precautions, I was affected by COVID-19 along with my 14-month-old baby; my*

XSL•FO
RenderX

> *baby has already suffered because of me. My husband also tested positive and was admitted to the hospital. We couldn't maintain proper isolation, as my baby was very young then. I wasn't seriously affected, but the baby was.* [Participant 3]

During the peak of the COVID-19 outbreak, the patient influx overwhelmed hospitals, pushing them beyond capacity. Therefore, all wards, cabins, emergency units, intensive care units, and critical care units were filled with susceptible patients with COVID-19. This created additional challenges for physicians trying to admit their sick relatives. Participant 6 recounted her struggle to admit her father to the hospital. Meanwhile, her sisters and mother were at home, infected with COVID-19, and their health was deteriorating, adding more to her stress:

> *I took my father to the hospital and asked for a cabin. There were no empty beds. I called my principal, who advised me to go to another hospital. It took me two hours as a doctor to find a space there; a random person would have needed much longer. My two sisters and my mother were also affected by COVID-19. They couldn't comprehend how my mother's oxygen saturation was falling. After three days, I went home, and there was no one else I could ask for help. I eventually took my mother to the cabin where my father was and provided her with oxygen. That night was incredibly stressful.* [Participant 6]

These experiences reflect a common theme among physicians, showcasing their declining mental and physical well-being as they navigated the dual responsibilities of caring for patients in the hospital and their families at home.

Theme 2: The Hidden Battles of COVID-19: Physicians' Burnout, Wellness, and Social Stigmatization

Navigating Physicians' Workload and Wellness

Physicians' pressures of their professional obligations profoundly impacted their personal lives, leading to multilayered experiences. During the COVID-19 pandemic, the work schedule for health care professionals was severely disrupted for 2 main reasons: an imbalance in the ratio of physicians to patients who were infected compared to normal times and the unavailability of coworkers due to COVID-19 infections. Participant 13 mentioned, "We, four doctors, are treating 50-60 patients simultaneously." Most patients with COVID-19 were admitted to the medicine unit, significantly increasing the workload. The physicians reported working closely with their senior colleagues to manage the high volume of patients, which led to behavioral changes and occasional outbursts among the staff. Participant 1 shared the following:

> *In the medicine unit, the mental pressure is greater than that of manual labor. Our seniors have to work even harder, which results in behavioral changes like bad tempers; some even misunderstand each other for no reason and shout.*

The rapidly deteriorating conditions of the patients added to the pressure, as physicians frequently had to revisit the same patients while simultaneously attending to new ones. Due to that, many experienced back-to-back morning and night shifts, which were more burdensome than usual. There were instances of 24-hour shifts, and after only a few hours of rest, the physicians had to return to work. This situation was exacerbated by the absence of many coworkers due to COVID-19 pandemic, forcing those present to take on additional responsibilities, negatively affecting their well-being. Participant 11 articulated the complexity of their duties and the resultant stress:

> *When co-workers skip their duty, we cannot skip ours. If they are assigned 12 patients, we ultimately have to manage their share, which adds pressure on us. It feels especially overwhelming to think about a week of consecutive night duties during that time. Compared to then, duty slots are better now.*

These demands greatly impacted their sleep quality. Many physicians experienced difficulties sleeping. Some developed insomnia during the pandemic, which took a long time to resolve. Participant 14 stated the following:

> *I suffered from insomnia for a while, but establishing a good routine helped me. During that time, I had to take sleeping pills.*

This indicates that when they desperately needed sleep, they found it difficult to rest without medication, further disrupting their sleep patterns. Participant 10 shared the following:

> *I sleep very little at night. I have developed insomnia, so I usually sleep around dawn.*

This lack of sleep adversely affected their physical health, leaving them feeling drowsy, with headaches and nausea, which in turn affected their ability to care for patients. Participant 13 remarked the following:

> *When I don't sleep enough, I feel drowsy. If my supervisor asks me about a patient's condition, I stare blankly because I didn't get adequate sleep; I feel nauseous and have a headache.*

This highlighted the lack of support toward the well-being of physicians during the intense period of the pandemic.

The Unseen Social Isolation and Stigmatization

Physicians experienced negative social perceptions during the pandemic due to their profession. Many people in their neighborhoods and relatives ignored them, seeing physicians as potential carriers of the virus because of their direct connection with hospitals. This perception took a toll on physicians' mental health, causing them to feel consistently down.

Many physicians struggled to manage their accommodation during this time because local residents were unwilling to live near them. Participant 3 mentioned that she had to secretly live in a place while hiding her medical laboratory coat:

> *We found a place through a relative, but we were required to take off our lab coats when returning home so that people wouldn't know we worked in a COVID-19 dedicated hospital.*

XSL•FO
RenderX

Therefore, physicians reported encountering negative attitudes from their neighbors, directly or indirectly. Participant 9 shared that no one in her building knew she was a physician:

> *Most people didn't know I am a doctor in my flat. If they knew, they might have treated me like others. I never wore my lab coat while going to work.*

Many physicians had to change their living arrangements multiple times. They initially stayed in hotels, but due to costs and safety concerns, they moved to a closed school where they had to sleep on the floor. However, the school committee eventually forced them to relocate. These challenges arose solely because of their work. In even worse situations, relatives refused to let them stay with them, and landlords denied rental applications from physicians. Participant 3 highlighted the mental pressure and uncertainty surrounding these issues:

> *No relatives would let us stay with them. We struggled to find rentals as landlords immediately refused to rent to us once they knew we were doctors.*

Consequently, during Ramadan in the COVID-19 period, the physicians did not have food facilities inside the hospital. So, they had to order food from outside, but they often faced negative attitudes from food delivery riders, who viewed hospitals as risky zones. Therefore, physicians frequently encountered rude behavior from these riders. Participant 6 stated this unhappily:

> *A rider asked me to come downstairs. I explained that I couldn't and asked him to come near the gate instead. He rudely replied, "We don't have any rules for that. COVID-19 is greater in the hospital." I responded, "Do we produce COVID-19 in medical facilities? If you or your family members were affected by COVID-19, wouldn't you come here? I have faced such dire situations several times."*

Physicians often faced harassment from traffic police while going out on duty. They had to show their identification to prove they were physicians, but frequently, the police did not believe them and prevented them from reaching their workplaces. The police prioritized enforcing nationwide lockdown restrictions over the physicians' essential roles in hospitals. Participant 2 expressed her frustration, saying:

> *When a doctor goes out and shows their ID, the police don't accept it. They say, "What's the guarantee that you have a duty now?" It's social harassment.*

This lack of support and social negligence was high during the initial period of the COVID-19 pandemic (March 2020-May 2020).

Theme 3: Emotional Toll and Lack of Mental Support on Physicians

Physicians often struggle to manage their emotions regarding their patients' conditions. They experience mixed feelings; they feel joy when patients recover but become stressed and emotionally exhausted when patients worsen. During the pandemic, however, many physicians found it increasingly difficult to cope with their emotions due to the rapid decline and mass deaths of patients. Constant sadness became a regular part of their experience, and some could not handle it. Participant 2 shared a painful memory:

> *She was 22 years old, with 90% lung problems, and I put her on high-flow oxygen. She had a 1-year-old child who was crying for milk beside her, but I couldn't stabilize his mother. I felt desperate to save her in any way possible. But eventually, she passed away. It was excruciating for me. Similarly, another patient was admitted two months ago with COVID-19 complications affecting his lungs. His wife was expecting their second baby. We tried desperately to save him, but suddenly, he passed away. These experiences affected us emotionally a lot.*

When physicians treat patients for an extended period, they often develop strong emotional attachments, especially to those older or similar in age. Patients can feel like family to them. The physicians invest a great deal of effort into trying to improve their patients' conditions, and when they are unable to do so, they experience significant emotional experiences. Participant 6 noted the following:

> *I stayed in a ward for so long that there was an attachment. It felt terrible seeing 28- to 30-year-old boys and girls come into the ward with COVID-19 and then witnessing their drastic decline. I couldn't do anything.*

Declaring a patient's death adds another layer of stress for physicians, as it involves both emotional and professional repercussions, particularly when informing the patient's family. Physicians experience immense psychological strain as they repeatedly witness their patients die. Participant 1 expressed her feelings:

> *The issue of declaring death causes my heart to race irregularly. One night shift, I was sad and had to declare four deaths in one night. I couldn't sleep that night; it was my most stressful night.*

Despite these overwhelming situations, many physicians reported a lack of mental support or counseling to help them manage the emotional toll. They often resort to self-counseling to cope, while in many Western countries, physicians' mental health is routinely prioritized. Participant 5 highlighted this gap in the Bangladeshi medical system, where taking a leave to focus on mental health is nearly impossible:

> *There is no counseling system. You have to continue your duties and manage everything on your own. Even if you must take leave for your mental well-being, you must find a replacement doctor to assume your duties. No matter what, you cannot leave the hospital ward empty. A co-worker of mine struggled greatly when her parents were hospitalized with COVID-19, but she couldn't take leave and had to fulfill her responsibilities.*

Physicians also encountered significant limitations in providing counseling services to patients' families during COVID-19. The sudden, tragic deaths of patients, who were often otherwise healthy, left families overwhelmed. Participant 2 commented the following:

XSL•FO
RenderX

> *Counseling services are limited in our country. When a patient passes away, their family members often don't want to hear anything. It's tough. Sometimes, one family member blames another for their own mental satisfaction.*

This underscores the need for counseling sessions to support mental health among all stakeholders within the health care system. Participant 14 suggested the following:

> *It would be very helpful if we had a group where doctors could share their problems and consult each other. Learning from others' experiences would help us mentally prepare for challenging cases and provide comfort in difficult situations.*

Discussion

Principal Findings

Our results highlighted the struggles faced by Bangladeshi physicians and the decline in their physical and mental well-being during the COVID-19 pandemic. These physicians had to attend to an excessive number of patients, often in close physical contact, exposing themselves to the risk of infection. This heightened their anxiety, especially with fears of transmitting the virus to their families. Many physicians, along with their family members, contracted COVID-19, which hindered their ability to provide adequate support at home due to the widespread crisis. In some cases, physicians chose not to reveal their COVID-19 symptoms to reduce fear and panic among their family members. As they worked tirelessly to manage a high volume of patients, they often could not take necessary breaks, even when unwell. This overwhelming workload compromised their physical health. Moreover, instead of receiving societal support, many physicians faced negative attitudes from the public, which further impacted their mental health. The loss of patients also took an emotional toll on these health care professionals. While counseling could help them cope with these challenges, the lack of a nationwide counseling system leaves many physicians feeling stressed and burned out.

To support physicians' mental and physical well-being, we need to consider well-accepted contextual support interventions that involve physicians and other stakeholders in society. During the pandemic, Daniels et al [31] proposed a grounded and coherent model based on the experiences and psychological care needs of frontline COVID-19 physicians in the United Kingdom, derived from ethnographic research. This credible model inspired us to develop our own contextual model, incorporating similar components while expanding to include available technological solutions for supporting physicians' health.

The Pyramid Model

The pyramid bottom-up model by Daniels et al [31] of well-being and psychological care for physicians consists of four layers: (1) basic needs and physiological resources, (2) information and communication, (3) embedded support, and (4) psychological interventions, as illustrated in Figure 1.

The first layer emphasized the importance of implementing local guidelines for proper nutrition, rest, and sleep. They suggested creating designated, quiet, confidential spaces for the physicians and the staff to reflect on and seek support. Fostering a culture of care and shared responsibility for staff well-being is essential, with senior leaders playing a key role in modeling positive behaviors related to well-being. In addition, they highlighted the need for personal safety measures.

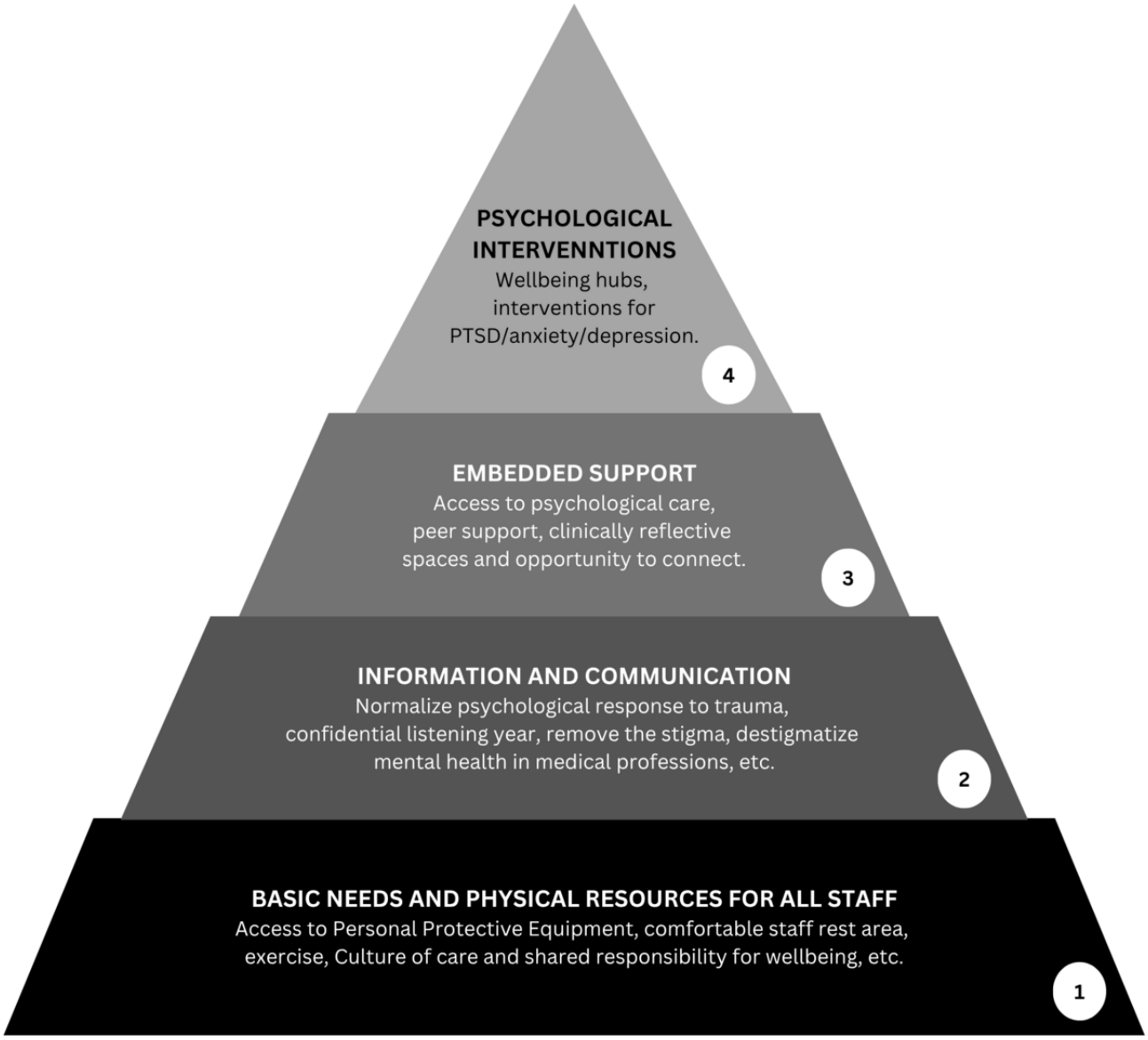
The second layer focused on the fact that stressful and traumatic incidents require open discussions on their psychological effects and the promotion of mental health awareness in medical training. Staff should be educated on warning signs of mental health deterioration, such as emotional responses and absenteeism, and should have a safe space to discuss their well-being without fear of negative repercussions.

The third layer emphasized that all physicians should have access to integrated psychological support. This includes face-to-face or video call reflective sessions among peers and senior staff within clinical teams. In addition, structured peer-to-peer interventions, such as Trauma Risk Management or the "StartWell>EndWell" approach, should be readily available for enhanced support.

Finally, in the last layer, psychological interventions should be tailored to individual needs while considering team well-being and available in group, personal, and online formats. It should be able to manage the impact of fear, anxiety, panic, guilt, hopelessness, disconnection, and compassion fatigue among health care workers. Support should address issues such as moral injury, trauma, posttraumatic stress disorder, depression, and mental health stigma.

XSL•FO
RenderX

Figure 1. Model of well-being and psychological care for frontline physicians (reproduced from Daniels et al [31], which is published under Creative Commons Attribution 4.0 International License [40]). PTSD: posttraumatic stress disorder.



Comparison Analysis

The research by Daniels et al [31] shows physicians often cope with stress through various strategies, such as exercise, spending time outdoors, expressing their feelings, practicing mindfulness, and journaling, which is absent for Bangladeshi physicians. However, many reported storing emotions due to limited time to process events, leading to emotional disconnection, particularly when dealing with patients who are severely ill or dying. Bangladeshi physicians also cannot vent out their emotions. Physicians in the United Kingdom often prioritized others' needs over their own, setting aside personal well-being to support colleagues or patients, which a few physicians in Bangladesh also do. The research by Daniels et al [31] shows support systems included talking to coworkers, accessing embedded psychological services, formal therapy, mental health apps, phone support, and messaging groups, as well as relying on friends and family. Barriers to support included lack of time for discussions or attending sessions. Workplace wellness initiatives were mentioned, with small acts of kindness or environmental adjustments having a notable positive impact. Negative influences included poor communication, ineffective leadership, and government decisions. However, in Bangladesh, no support system exists for the physicians to assist them in their well-being. Hence, an improvement in the system is required along with a supportive network. This motivates us to work for the contextual support system, influenced by the research model by Daniels et al [31].

Proposed Trajectile Model for Physicians' Well-Being

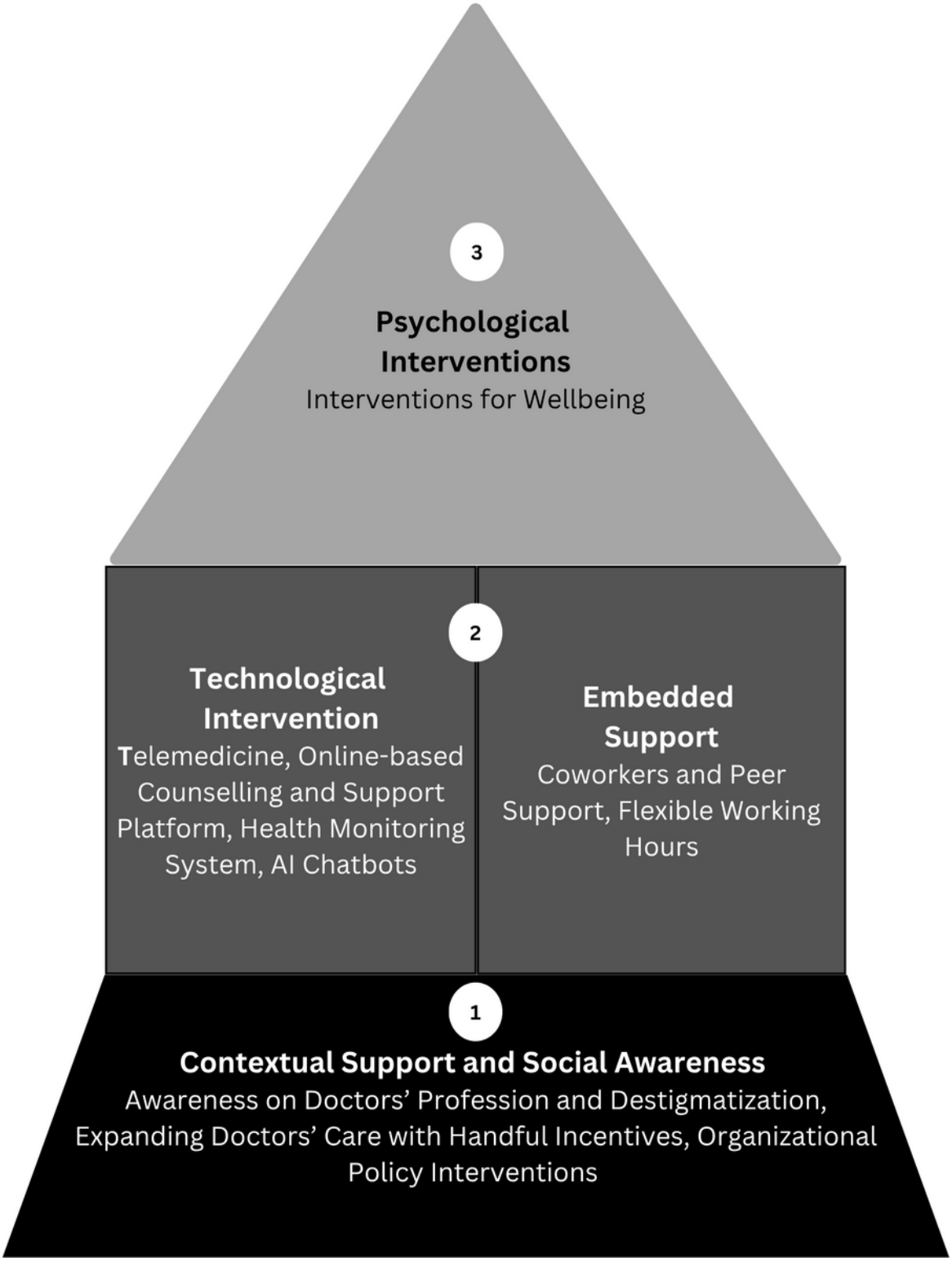
Overview

Our trajectile bottom-up model for physicians' well-being is depicted in 3 layers, shown in Figure 2. We built upon our findings to develop this model and incorporated contextual and technological components, including AI interventions. The layers of our model include (1) contextual support and social awareness, (2) technological interventions and embedded support, and (3) psychological interventions. We aligned our findings with the proposed model by Daniels et al [31] (Figure 1) and then modified and integrated components to create our

XSL•FO
RenderX

proposed model (Figure 2). For example, the embedded support and psychological intervention layers are similar to the model by Daniels et al [31]. We recommend contextual support and social awareness, and the technological interventions components in the trajectile model differ from the model by Daniels et al [31], but that inspired us. The proposed model can be a sustainable model with effective implementation and can be adaptable in any situation and culture.

Figure 2. Proposed trajectile model for physicians' well-being. AI: artificial intelligence.



XSL•FO
RenderX

Layer 1: Contextual Support and Social Awareness

Overview

In the first layer of the trajectile models, we recommend enhancing contextual support and social awareness for physicians. This includes raising awareness about the medical profession, reducing stigma, expanding support for physicians through various incentives, and implementing policy interventions aimed at promoting both their mental and physical well-being.

Awareness of Physicians' Profession and Destigmatization

We found that physicians often faced negative attitudes and stigma from society during the COVID-19 pandemic, which can lead to feelings of stress. In addition, physicians tend to refrain from expressing their emotions in the workplace because they fear others might interpret their feelings as signs of fear or guilt rather than understanding that they genuinely care for their patients (referring to themes 2 and 3). In Bangladesh, for instance, many people blame physicians for patient losses due to perceived negligence [41]. Accordingly, Galbraith et al [13] highlighted methods for destigmatizing the workplace for physicians, such as fostering an open communication culture, organizing team-building activities, and providing peer support training during health crises and traumatic events, which can include nonmedical personnel [42,43].

However, in our context, most stigmatization occurs at the societal level. It is crucial for society to understand the challenges and hardships that physicians face and to offer expanded support accordingly. Key strategies for preventing blame and psychological trauma among physicians include raising public awareness, allowing physicians to maintain a sense of self-determination, and encouraging effective communication skills [28]. By implementing these strategies with the stakeholders around physicians, society can better appreciate physicians' dedication to their patient care.

Expanding Physicians' Care With Handful Incentives

Our physicians have mentioned that they often feel outnumbered by the increasing number of patients during the pandemic (referring to theme 2). This situation highlights the need to increase the number of physicians [44-47]. A significant reason is that rural patients tend to migrate to cities because there are fewer physicians in rural areas. In addition, many physicians are reluctant to work in these regions due to logistical challenges, such as low incentives, limited medical support, and inadequate accommodation options [48]. Previous research has also indicated that physicians do not receive sufficient recognition through incentives [21,28,44,47]. Bangladesh faces a similar issue, with only 5.26 physicians available for every 10,000 people [48]. Furthermore, their incentives are minimal, often no more than US $270, largely due to bureaucratic obstacles [49]. To address the shortage of physicians, the government should expand medical education and specialized training programs to ensure that more physicians are prepared to serve the rural population effectively. Establishing policies that provide adequate incentives in these areas is also essential. Many countries have already adopted performance-based incentive programs alongside salaries, and evidence shows that such incentives can significantly improve health care providers' performance [50-52]. By implementing these measures, we can develop a nationwide health care system that is more evenly distributed and sustainable. This system will help alleviate the patient overload in urban areas while extending vital medical support to rural regions.

Organizational Policy Interventions

Physicians often work irregular hours, which negatively impacts their well-being (referring to theme 2). To address this issue, hospital administration must revise policies to support physicians' health for the greater good. Certain measures, such as providing personal space for rest, protecting against occupational hazards, offering childcare opportunities, and implementing incentives and maternity and parental leave, can help prevent burnout among health care workers [31,53]. However, many public hospitals in Bangladesh lack these provisions, while some private hospitals offer them only to a limited number of physicians. Therefore, hospital administration should implement these measures within the health care system so the government can take the necessary steps to mandate these policies.

Layer 2: Technological Intervention and Embedded Support

In this second layer, we included 2 components: technological interventions (such as telemedicine, online support platforms, health monitoring systems, and AI chatbots) and embedded support (such as coworkers and peer support and flexible working hours), essential for improving physicians' health.

Technological Intervention

Telemedicine

Our physicians have expressed concerns about the risk of COVID-19 infection from undiagnosed patients and the necessity of physical contact (referring to theme 1). Telemedicine is crucial in reducing physical interactions between physicians and patients, thereby minimizing the risk of exposure to various viruses and diseases, including the SARS-CoV-2 [54]. Physicians can provide clinical care through telephone and video consultations, especially since COVID-19 significantly affects older populations [55].

Bangladesh offered several telemedicine services during the pandemic, such as "Tonic" and "ShebaGhor." However, these services struggled to sustain themselves due to inadequate regulations, patient trust, and limited resources [54]. In addition, many people in Bangladesh are not accustomed to telemedicine and face challenges related to access to information and communication technology, leading them to prefer in-person consultations. Nevertheless, we advocate for the adoption of telemedicine, as general practice in the United Kingdom has successfully transitioned to telephone or video consultations. This shift has reduced unnecessary face-to-face interactions and the associated risk of virus transmission, easing concerns about contracting illnesses [56].

The effectiveness of telemedicine is evident in high-income nations, where the system is regulated by the government. In contrast, telemedicine services in Bangladesh operate

XSL•FO
RenderX

independently, raising public trust issues. The Bangladesh Medical and Dental Council should establish proper regulations to create a sustainable telemedicine system. Hospitals could then operate toll-free telemedicine hotlines to broaden services and enhance accessibility. A single toll-free number could connect callers to local hospitals. Bangladesh already has emergency service numbers in place; using and expanding this existing infrastructure could help address the needs of a large patient population and benefit physicians' health.

Online Counseling and Support Platform

Physicians often seek counseling to express their emotions, but they face a lack of options, particularly in resource-limited countries such as Bangladesh, where hospitals do not have proper counseling units (referring to theme 3). Digital platforms can support physicians' mental health by providing easy access to counseling services, peer connections, and stress management tools. These resources can help them navigate the challenges of their profession and alleviate stress. Digital and technological platforms supplement traditional behavioral support [57]. Counseling delivered over the phone or through the internet can effectively manage physicians' mental health [58], but it should be easily available and accessible and timely [59,60]. To address this need, the Bangladesh Medical and Dental Council could establish an online counseling platform accessible only to physicians, enabling them to receive the support they require from anywhere.

Health Monitoring System

We additionally recommend technology-based health monitoring systems that can assist physicians in tracking their health and mental well-being. As participants are lacking in maintaining their personal well-being at this point, cost-effective wearable devices, such as Fitbit (Google LLC), offer valuable insights into various health metrics and self-assessment. For instance, Fitbit can monitor sleep patterns, physical activity, blood oxygen levels, heart rate variability, and exercise frequency, all of which can indicate a person's healthy lifestyle [61] and stress levels. This enables physicians to support their own health and pursue a healthier lifestyle. In addition, Anmella et al [62] developed 2 chatbots designed to screen, monitor, and mitigate anxiety and depressive symptoms as well as work-related burnout while also assessing suicide risk in physicians through conversational interactions. Such technology can help identify stress among physicians and alert relevant authorities or family members, enabling them to support the physicians' well-being. Meanwhile, physicians can use these tools to engage in self-care and self-medication therapies.

AI Chatbots

There is often a significant lack of resources in resource-limited countries, leading to accessibility issues for counseling tools and digital platforms. Participants also explicitly mentioned their lack of counseling support (referring to theme 3). AI chatbots can provide a cost-effective solution for resource-limited countries such as Bangladesh. Baek et al [63] suggested that AI chatbots can help improve states of anxiety, depression, and burnout. For instance, the Woebot chatbot has assisted young adults with depression and anxiety, resulting in a significant reduction in both conditions [64]. Similarly, the Vivibot chatbot contributed to reducing anxiety, although it did not have an effect on depression [65]. Today, AI technologies, such as ChatGPT by OpenAI, Gemini by Google LLC, Gork by X, and Meta from Meta are widely accessible and can offer open-ended well-being suggestions. However, these systems often lack personalization, making it difficult to address specific mental health issues effectively. Designing a mobile app that combines existing chatbots while incorporating personalized features is feasible. Therefore, AI chatbots could serve as a valuable tool to support the psychological well-being of physicians by providing personalized and accessible interventions, especially in resource-limited countries.

Embedded Support

Coworkers and Peer Support

We observed that when coworkers fall ill and request assistance, they must address those issues immediately, increasing pressure during the pandemic (referring to theme 2). On the other hand, coworkers also supported one another during their illnesses. This support is essential in the medical field to maintain workplace efficiency and the mental well-being of physicians. Having reliable, responsible, empathetic, accountable, and supportive coworkers during the pandemic was invaluable for navigating challenges. Senior coworkers provided clear guidelines, served as role models for junior physicians, and fostered a culture of learning and collaboration among peers [22]. This support is also considered peer support, which can be extended to social support with regular contact via phone call or video chat with colleagues, family, and friends [57]. Therefore, these supportive relationships can significantly enhance one another's well-being.

Flexible Working Hours

Physicians have said that they often work long hours without any breaks, leading to mental exhaustion (referring to theme 2). They should be given a flexible schedule with adequate breaks between shifts to address this issue. In Bangladesh, interns experience unstructured duty hours; most physicians work daily. Urooj et al [22] suggest that minimizing physicians' duty hours and limiting routine checkups can improve their health and reduce the number of staff working at any given time, allowing some to be reserved as backups. This approach would help achieve a better work-life balance [53]. Therefore, the administration needs to redefine and structure physicians' working hours, and the Bangladesh Medical and Dental Council should implement a policy to support these changes. This would ensure that the support provided is appropriately fixed and effective.

Layer 3: Psychological Interventions

We refer to psychological interventions similar to the model proposed by Daniels et al [31], which focuses on addressing individuals' needs while managing the effects of anxiety, hopelessness, and stress. The physicians involved in the study expressed a desire for regular monthly counseling sessions due to the significant loss of patients during every crisis. Therefore, counselors and volunteers could play a crucial role in not only addressing acute symptoms, such as depression, anxiety, and psychological distress, but also in supporting physicians during

XSL•FO
RenderX

the postpandemic phase and helping them cope with ongoing challenges. This support intervention should be available online and in hospitals with easy access.

Limitations and the Way Forward

We interviewed 14 participants online during the pandemic, all from the Chattogram region. This region is around 300 km southeast of the capital but is an overpopulated city. The medical support system was weaker than the capital's. We were able to access these physicians during this emergency period. Although we aimed to expand our interviews to different cities, such as the capital Dhaka, so that we could perform comparative analysis, many potential participants did not respond in Dhaka. This presents a limitation, as the findings may not be generalizable. However, they reflect physicians' overall experiences during the pandemic, which aligns with the situation of most physicians who worked during the COVID-19 pandemic in Bangladesh. In addition, the proposed model was not tested or validated. Testing the model with the stakeholders may give us valuable insight into its acceptability and usability.

Future research can be expanded by, including physicians and stakeholders from hospitals and the Bangladesh Medical and Dental Council, as many institutional issues are linked to hospital administration and policy makers within the council. Engaging these groups in further studies will provide insights into effective interventions for maintaining physicians' well-being. In addition, we plan to conduct longitudinal studies using wearable technologies, such as Fitbits, to gather real-time physiological data from physicians. Analyzing these data will help identify stress factors associated with their work, which may contribute to redesigning their work environment in the future.

Conclusions

A qualitative study involving 14 physicians explored their experiences and well-being during the COVID-19 pandemic. This study helped us identify their challenges in maintaining their well-being amid various contextual difficulties. We recognized the importance of addressing these concerns and set out to develop a contextual intervention model. Our model for promoting well-being was inspired by the well-being framework by Daniels et al [31], incorporating contextual and technological interventions. We proposed suggestions based on a bottom-up approach, highlighting interconnected concepts. By focusing on this model, we aim to engage stakeholders in rethinking and redesigning the work environment for physicians. This approach is designed to be sustainable over the long term and adaptable to any situation and culture.

Acknowledgments

The authors thank all their participants for their cooperation and support. This study was supported by the Design Inclusion and Access Lab at North South University, along with small seed funding (for buying equipment and participant incentives) from the University of British Columbia. The authors extend their special thanks to Md Sabbir Ahmed from the University of Virginia and Mohammad Hadi from the University of British Columbia for their contributions during this study's design and interview phases.

Authors' Contributions

RJR and NA designed and conducted the qualitative study, with NA supervising the overall process. RJR played a significant role in writing the manuscript from start to finish, as well as supervising the data analysis. SAA, LAT, and MM contributed to the data analysis, conducted the literature review, and worked on the introduction and discussion sections. All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Polizzi A, Zhang L, Timonin S, Gupta A, Dowd JB, Leon DA, et al. Indirect effects of the COVID-19 pandemic: a cause-of-death analysis of life expectancy changes in 24 countries, 2015 to 2022. PNAS Nexus 2025 Jan;4(1):pgae508 [FREE Full text] [doi: 10.1093/pnasnexus/pgae508] [Medline: 39720204]
2. Zaidi Z. A statistical review of the impact of COVID-19 on the world. Transdiscipl J Eng Sci 2024 Apr 09;15:93 [FREE Full text] [doi: 10.22545/2024/00251]
3. Cawley C, Barsbay M, Djamangulova T, Erdenebat B, Cilović-Lagarija Š, Fedorchenko V, BoCO-19-Study Group. The mortality burden related to COVID-19 in 2020 and 2021 - years of life lost and excess mortality in 13 countries and sub-national regions in Southern and Eastern Europe, and Central Asia. Front Public Health 2024;12:1378229 [FREE Full text] [doi: 10.3389/fpubh.2024.1378229] [Medline: 38903591]
4. Rousson V, Locatelli I. Years of life lost to COVID-19 and related mortality indicators: an illustration in 30 countries. Biom J 2024 Jul;66(5):e202300386 [FREE Full text] [doi: 10.1002/bimj.202300386] [Medline: 39001703]
5. Iyengar KP, Ish P, Upadhyaya GK, Malhotra N, Vaishya R, Jain VK. COVID-19 and mortality in doctors. Diabetes Metab Syndr 2020;14(6):1743-1746 [FREE Full text] [doi: 10.1016/j.dsx.2020.09.003] [Medline: 32920494]

XSL•FO
RenderX

6. Hummel S, Oetjen N, Du J, Posenato E, Resende de Almeida RM, Losada R, et al. Mental health among medical professionals during the COVID-19 pandemic in eight European countries: cross-sectional survey study. J Med Internet Res 2021 Jan 18;23(1):e24983 [FREE Full text] [doi: 10.2196/24983] [Medline: 33411670]
7. Ahmari AS, Alqahtani AS, Asiri FH, Abdullah Sulaiman Samiti F, Alqahtani KA. Balancing duty and risk: a critical analysis of frontline medical workers during health emergencies. Journal of Ecohumanism 2024;3(8):9034-9041 [FREE Full text] [doi: 10.62754/joe.v3i8.5520]
8. Berg S. New study tallies excess physician deaths during early pandemic. American Medical Association. 2023 Feb 07. URL: https://www.ama-assn.org/practice-management/physician-health/new-study-tallies-excess-physician-deaths-during-early [accessed 2025-05-29]
9. COVID-19 deaths in doctors and nurses, 2020 to 2022. Office for National Statistics. URL: https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/covid19deathsindoctorsandnurses2020to2022 [accessed 2025-05-29]
10. Purohit J. Covid-19: India health workers' families fight for compensation. BBC. URL: https://www.bbc.com/news/world-asia-india-58621933 [accessed 2025-05-29]
11. Mao W, Chen M. Timely sharing of data on infection and death of medical workers. Front Public Health 2020;8:552409 [FREE Full text] [doi: 10.3389/fpubh.2020.552409] [Medline: 33194953]
12. Rahman MD, Farzana M, Rahman T. Occupational challenges and physician deaths during COVID-19 pandemic. Int J Travel Med Glob Health 2021 Jan 06;9(2):100-101 [FREE Full text] [doi: 10.34172/ijtmgh.2021.16]
13. Galbraith N, Boyda D, McFeeters D, Hassan T. The mental health of doctors during the COVID-19 pandemic. BJPsych Bull 2021 Apr;45(2):93-97 [FREE Full text] [doi: 10.1192/bjb.2020.44] [Medline: 32340645]
14. Sun H, Zhang T, Wang X, Wang C, Zhang M, Song H. The occupational burnout among medical staff with high workloads after the COVID-19 and its association with anxiety and depression. Front Public Health 2023;11:1270634 [FREE Full text] [doi: 10.3389/fpubh.2023.1270634] [Medline: 37954047]
15. Preti E, Di Mattei V, Perego G, Ferrari F, Mazzetti M, Taranto P, et al. The psychological impact of epidemic and pandemic outbreaks on healthcare workers: rapid review of the evidence. Curr Psychiatry Rep 2020 Jul 10;22(8):43 [FREE Full text] [doi: 10.1007/s11920-020-01166-z] [Medline: 32651717]
16. Ghahramani S, Lankarani KB, Yousefi M, Heydari K, Shahabi S, Azmand S. A systematic review and meta-analysis of burnout among healthcare workers during COVID-19. Front Psychiatry 2021;12:758849 [FREE Full text] [doi: 10.3389/fpsyt.2021.758849] [Medline: 34858231]
17. Dwivedi M, Pandey S. Stress among doctors during COVID-19. Int J Indian Psychol 2020;3:63 [FREE Full text] [doi: 10.25215/0802.230]
18. Elbay RY, Kurtulmuş A, Arpacıoğlu S, Karadere E. Depression, anxiety, stress levels of physicians and associated factors in COVID-19 pandemics. Psychiatry Res 2020 Aug;290:113130 [FREE Full text] [doi: 10.1016/j.psychres.2020.113130] [Medline: 32497969]
19. Doleman G, De Leo A, Bloxsome D. The impact of pandemics on healthcare providers' workloads: a scoping review. J Adv Nurs 2023 Dec;79(12):4434-4454 [FREE Full text] [doi: 10.1111/jan.15690] [Medline: 37203285]
20. Razu SR, Yasmin T, Arif TB, Islam MS, Islam SM, Gesesew HA, et al. Challenges faced by healthcare professionals during the COVID-19 pandemic: a qualitative inquiry from Bangladesh. Front Public Health 2021;9:647315 [FREE Full text] [doi: 10.3389/fpubh.2021.647315] [Medline: 34447734]
21. Mohd Fauzi MF, Mohd Yusoff H, Muhamad Robat R, Mat Saruan NA, Ismail K, Mohd Haris AF. Doctors' mental health in the midst of COVID-19 pandemic: the roles of work demands and recovery experiences. Int J Environ Res Public Health 2020 Oct 08;17(19):7340 [FREE Full text] [doi: 10.3390/ijerph17197340] [Medline: 33050004]
22. Urooj U, Ansari A, Siraj A, Khan S, Tariq H. Expectations, fears and perceptions of doctors during COVID-19 pandemic. Pak J Med Sci 2020 May 18;36(COVID19-S4):19-S4 [FREE Full text] [doi: 10.12669/pjms.36.covid19-s4.2643]
23. Emre N, Edirne T, Ozsahin A, Kulceler MF. Assessment on risk and stress of resident doctors during the COVID-19 pandemic. J Infect Dev Ctries 2021 Aug 31;15(8):1080-1085 [FREE Full text] [doi: 10.3855/jidc.14877] [Medline: 34516414]
24. Khatun MF, Parvin MF, Rashid MM, Alam MS, Kamrunnahar M, Talukder A, et al. Mental health of physicians during COVID-19 outbreak in Bangladesh: a web-based cross-sectional survey. Front Public Health 2021;9:592058 [FREE Full text] [doi: 10.3389/fpubh.2021.592058] [Medline: 33634065]
25. Babamiri M, Alipour N, Heidarimoghadam R. Research on reducing burnout in health care workers in critical situations such as the COVID-19 outbreak. Work 2020;66(2):379-380. [doi: 10.3233/WOR-203189] [Medline: 32568163]
26. Amin F, Sharif S, Saeed R, Durrani N, Jilani D. COVID-19 pandemic- knowledge, perception, anxiety and depression among frontline doctors of Pakistan. BMC Psychiatry 2020 Sep 23;20(1):459 [FREE Full text] [doi: 10.1186/s12888-020-02864-x] [Medline: 32967647]
27. Kinman G, Grant C. Presenteeism during the COVID-19 pandemic: risks and solutions. Occup Med (Lond) 2021 Oct 01;71(6-7):243-244 [FREE Full text] [doi: 10.1093/occmed/kqaa193] [Medline: 33205200]
28. Shrestha RM, Kunwar AR. COVID-19 impact on doctors and health workers. Orthod J Nepal 2020 Sep 11;10(2):2-5 [FREE Full text] [doi: 10.3126/ojn.v10i2.31060]

XSL•FO
RenderX

29. Zhang J, Lu X, Jin Y, Zheng ZJ. Hospitals' responsibility in response to the threat of infectious disease outbreak in the context of the coronavirus disease 2019 (COVID-19) pandemic: implications for low- and middle-income countries. Glob Health J 2020 Dec;4(4):113-117 [FREE Full text] [doi: 10.1016/j.glohj.2020.11.005] [Medline: 33294250]
30. Dutta U, Sachan A, Premkumar M, Gupta T, Sahoo S, Grover S, et al. Saving and supporting health care workers during the COVID-19 pandemic in a developing country using a multidimensional healthcare personnel centric policy. SSRN Journal. Preprint posted online December 7, 2023 2023 [FREE Full text] [doi: 10.2139/ssrn.3670616]
31. Daniels J, Ingram J, Pease A, Wainwright E, Beckett K, Iyadurai L, et al. The COVID-19 clinician cohort (CoCCo) study: empirically grounded recommendations for forward-facing psychological care of frontline doctors. Int J Environ Res Public Health 2021 Sep 14;18(18):632 [FREE Full text] [doi: 10.3390/ijerph18189675] [Medline: 34574598]
32. Frankenberg R. On unsteady ground: crafting and engaging in the critical study of whiteness. In: Bulmer M, Solomos J, editors. Researching Race and Racism. London, UK: Routledge; 2004:104-118.
33. Etikan I. Comparison of convenience sampling and purposive sampling. Am J Theor Appl Stat 2016;5(1):1. [doi: 10.11648/j.ajtas.20160501.11]
34. Braun V, Clarke V. Reflecting on reflexive thematic analysis. Qual Res Sport Exerc Health 2019 Jun 13;11(4):589-597 [FREE Full text] [doi: 10.1080/2159676x.2019.1628806]
35. Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. Qual Quant 2021 Jun 26;56(3):1391-1412 [FREE Full text] [doi: 10.1007/s11135-021-01182-y]
36. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115 [FREE Full text] [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
37. Glaser BG, Strauss AL. Discovery of Grounded Theory: Strategies for Qualitative Research. New York, NY: Routledge; 1967.
38. Hossain P, Das Gupta R, YarZar P, Salieu Jalloh M, Tasnim N, Afrin A, et al. 'Feminization' of physician workforce in Bangladesh, underlying factors and implications for health system: insights from a mixed-methods study. PLoS One 2019;14(1):e0210820 [FREE Full text] [doi: 10.1371/journal.pone.0210820] [Medline: 30633775]
39. Mone FH, Ashrafi DM, Sarker MA. Work life balance of female doctors in Bangladesh: an overview. J Health Med Sci 2019;2(3):5 [FREE Full text] [doi: 10.31014/aior.1994.02.03.65]
40. Attribution 4.0 International (CC BY 4.0). Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/
41. Shahjalal MD, Alam MM, Khan MD, Sultana A, Zaman S, Hossain A, et al. Prevalence and determinants of physical violence against doctors in Bangladeshi tertiary care hospitals. Hum Resour Health 2023 Mar 28;21(1):26 [FREE Full text] [doi: 10.1186/s12960-023-00811-x] [Medline: 36978102]
42. Halpern J, Gurevich M, Schwartz B, Brazeau P. Interventions for critical incident stress in emergency medical services: a qualitative study. Stress Health 2008 Oct 09;25(2):139-149 [FREE Full text] [doi: 10.1002/smi.1230]
43. Hanisch SE, Twomey CD, Szeto AC, Birner UW, Nowak D, Sabariego C. The effectiveness of interventions targeting the stigma of mental illness at the workplace: a systematic review. BMC Psychiatry 2016 Jan 06;16:1 [FREE Full text] [doi: 10.1186/s12888-015-0706-4] [Medline: 26739960]
44. Shahen M, Ahmed R. Challenges for health care services in Bangladesh: an overview. IOSR J Nurs Health Sci 2020;5:13-24 [FREE Full text] [doi: 10.9790/1959-0901011324]
45. Grimm C. Hospital experiences responding to the COVID-19 pandemic: results of a national pulse survey. Office of Inspector General. URL: https://justthenews.com/sites/default/files/2020-04/HHS2020ReportPandemicHospitals.pdf [accessed 2025-05-29]
46. Iacobucci G. Covid-19: UK needs more doctors to support post-pandemic recovery, say leaders. BMJ 2021 Oct 20;375:n2572 [FREE Full text] [doi: 10.1136/bmj.n2572] [Medline: 34670773]
47. Kruk ME, Prescott MR, de Pinho H, Galea S. Are doctors and nurses associated with coverage of essential health services in developing countries? A cross-sectional study. Hum Resour Health 2009 Mar 31;7:27 [FREE Full text] [doi: 10.1186/1478-4491-7-27] [Medline: 19335911]
48. Mohiuddin AK. An extensive review of patient health-care service satisfaction in Bangladesh. Patient Exp J 2020 Jul 23;2:5-16 [FREE Full text] [doi: 10.25259/AUJMSR_6_2020]
49. Bashar F, Islam R, Khan SM, Hossain S, Sikder AA, Yusuf SS, et al. Making doctors stay: Rethinking doctor retention policy in a contracted-out primary healthcare setting in urban Bangladesh. PLoS One 2022;17(1):e0262358 [FREE Full text] [doi: 10.1371/journal.pone.0262358] [Medline: 34986200]
50. Talukder M, Rob U, Rahman L, Hena I. Innovative financing through pay-for-performance for providers to improve quality of care in Bangladesh: Transforming research into action. Population Council. 2011. URL: https://doi.org/10.31899/rh2.1020 [accessed 2025-05-29]
51. Gesesew HA, Tebeje B, Alemseged F, Beyene W. Health workforce acquisition, retention and turnover in southwest Ethiopian health institutions. Ethiop J Health Sci 2016 Jul;26(4):331-340 [FREE Full text] [doi: 10.4314/ejhs.v26i4.5] [Medline: 27587931]
52. Ferry E, Retno I, Kurniati A. Incentives for retaining and motivating health workers in very remote areas of Indonesia. Southern Cross University. 2012. URL: https://doi.org/10.5176/2251-3833_ghc12.71 [accessed 2025-05-29]

XSL•FO
RenderX

53. Kumar S. Burnout and doctors: prevalence, prevention and intervention. Healthcare (Basel) 2016 Jun 30;4(3):37 [FREE Full text] [doi: 10.3390/healthcare4030037] [Medline: 27417625]
54. Khan MM, Rahman SM, AnjumIslam ST. The use of telemedicine in Bangladesh during COVID-19 pandemic. Ehealth Telecommun Syst Netw 2021;10(01):1-19 [FREE Full text] [doi: 10.4236/etsn.2021.101001]
55. Majeed A, Molokhia M, Pankhania B, Asanati K. Protecting the health of doctors during the COVID-19 pandemic. Br J Gen Pract 2020 May 11;70(695):268-269 [FREE Full text] [doi: 10.3399/bjgp20x709925]
56. Zhou Y, Wang L, Ding L, Tang Z. Intelligent technologies help operating mobile cabin hospitals effectively cope with COVID-19. Front Eng Manag 2020 Apr 21;7(3):459-460 [FREE Full text] [doi: 10.1007/S42524-020-0113-5]
57. Wong AH, Pacella-LaBarbara ML, Ray JM, Ranney ML, Chang BP. Healing the healer: protecting emergency health care workers' mental health during COVID-19. Ann Emerg Med 2020 Oct;76(4):379-384 [FREE Full text] [doi: 10.1016/j.annemergmed.2020.04.041] [Medline: 32534830]
58. Drissi N, Ouhbi S, Marques G, de la Torre Díez I, Ghogho M, Janati Idrissi MA. A systematic literature review on e-mental health solutions to assist health care workers during COVID-19. Telemed J E Health 2021 Jun;27(6):594-602 [FREE Full text] [doi: 10.1089/tmj.2020.0287] [Medline: 32970532]
59. Paice E, Moss F, West G, Grant J. Association of use of a log book and experience as a preregistration house officer: interview survey. BMJ 1997 Jan 18;314(7075):213-215 [FREE Full text] [doi: 10.1136/bmj.314.7075.213] [Medline: 9022443]
60. Afzal A, Kamal M, Diyanti N, Dellanira PR, Jaan A. Implementation of mandatory counseling sessions and availability of support centers for mental well being of heath care workers during the COVID-19 pandemic. Arch Intern Med Res 2021;04(01):33 [FREE Full text] [doi: 10.26502/aimr.0053]
61. Ringeval M, Wagner G, Denford J, Paré G, Kitsiou S. Fitbit-based interventions for healthy lifestyle outcomes: systematic review and meta-analysis. J Med Internet Res 2020 Oct 12;22(10):e23954 [FREE Full text] [doi: 10.2196/23954] [Medline: 33044175]
62. Anmella G, Sanabra M, Primé-tous M, Segú X, Cavero M, Navinés R, et al. Vickybot, a chatbot for anxiety-depressive symptoms and work-related burnout. Eur Psychiatr 2023 Jul 19;66(S1):S109-S110 [FREE Full text] [doi: 10.1192/j.eurpsy.2023.301]
63. Baek G, Cha C, Han JH. AI chatbots for psychological health for health professionals: scoping review. JMIR Hum Factors 2025 Mar 19;12:e67682 [FREE Full text] [doi: 10.2196/67682] [Medline: 40106346]
64. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JMIR Ment Health 2017 Jun 06;4(2):e19 [FREE Full text] [doi: 10.2196/mental.7785] [Medline: 28588005]
65. Greer S, Ramo D, Chang YJ, Fu M, Moskowitz J, Haritatos J. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: randomized controlled feasibility trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]

Abbreviations

AI: artificial intelligence

Edited by T de Azevedo Cardoso; submitted 04.05.25; peer-reviewed by D Lim, J Abua; comments to author 28.05.25; revised version received 07.07.25; accepted 14.07.25; published 26.09.25.

Please cite as:
Rony RJ, Aalok SA, Tisha LA, Mahatab M, Ahmed N
Understanding the Mental and Physical Burdens of Physicians and Identifying Support Interventions in Bangladesh: Qualitative Study
Interact J Med Res 2025;14:e76934
URL: https://www.i-jmr.org/2025/1/e76934
doi:10.2196/76934
PMID:41004260

©Rahat Jahangir Rony, Shams Akbar Aalok, Lamia Amin Tisha, Marzan Mahatab, Nova Ahmed. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.09.2025. 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

Compliance With Bowel Preparation and Its Influencing Factors in Patients Undergoing Colonoscopy: Cross-Sectional Study

Huan Jiang[1,2], MM; Chuanhui Li[1], BSN; Bing Hu[1,2], MD; Yi Mou[1,2], MD

[1]Department of Gastroenterology and Hepatology, West China Hospital, Sichuan University, 37 Guo Xue Alley, Wuhou District, Chengdu, China
[2]Digestive Endoscopy Medical Engineering Research Laboratory, West China Hospital, Sichuan University, China

Corresponding Author:
Yi Mou, MD
Department of Gastroenterology and Hepatology, West China Hospital, Sichuan University, 37 Guo Xue Alley, Wuhou District, Chengdu, China

Abstract

Background: Bowel preparation compliance is an important intervenable factor that affects bowel preparation quality, and improving compliance is an important way to optimize bowel preparation outcomes. Despite its importance, the compliance rate and its influencing factors have not been thoroughly evaluated.

Objective: This study aimed to investigate the overall compliance with bowel preparation instructions in patients undergoing colonoscopy.

Methods: From September 2024 to March 2025, a cross-sectional questionnaire-based study was conducted at West China Hospital of Sichuan University, recruiting 740 participants via convenience sampling. We used an 8-item self-report scale to evaluate compliance with bowel preparation instructions. Items were rated on a 4-point Likert scale (0=completely noncompliant to 3=completely compliant), yielding a total score of 0‐24. Higher scores reflected greater compliance, with ≥95% of the maximum score considered adequate compliance. Univariate analysis and multivariate logistic regression analysis were used to assess factors (age, educational level, knowledge of bowel preparation, satisfaction with the taste of the laxative, physical discomfort during bowel preparation) influencing bowel preparation compliance.

Results: In this study, 42.0% (311/740) of patients demonstrated adequate compliance with bowel preparation instructions. In the univariate analysis, hypertension history, knowledge of bowel preparation, laxative type, satisfaction with the taste of the laxative, anxiety during bowel preparation, and physical discomfort during bowel preparation all had statistically significant influences. Multivariate analysis showed that older age (odds ratio [OR] 2.27, 95% CI 1.16-4.49), higher educational level (OR 3.29, 95% CI 1.41-8.33), adequate knowledge of bowel preparation (OR 1.59, 95% CI 1.14-2.24), satisfaction with the taste of the laxative (OR 2.11, 95% CI 1.48-3.02), and no physical discomfort during bowel preparation (OR 0.45, 95% CI 0.31-0.64) were key factors for adequate bowel preparation compliance.

Conclusions: Personalizing bowel preparation instructions according to patients' age and education level, and selecting a laxative that suits the patients' taste preferences when available, are feasible ways to improve compliance with bowel preparation.

(Interact J Med Res 2025;14:e77189) doi:10.2196/77189

KEYWORDS

colonoscopy; bowel preparation; bowel preparation quality; compliance; adherence; influencing factors; survey; questionnaire; scale; cross-sectional study

Introduction

As the "gold standard" for the diagnosis of intestinal diseases such as colorectal cancer (CRC), colonoscopy provides direct and accurate visualization of the colonic mucosal surface, enabling early detection, accurate diagnosis, and timely intervention, thereby improving patient outcomes [1-3]. In 2021, the US Preventive Services Task Force recommended that people with risk factors such as a family history of CRC, obesity, inflammatory bowel disease, and smoking should undergo regular colonoscopies [4,5]. Even those without these risk factors should undergo CRC screening starting at the age of 45 years and have a colonoscopy every 10 years [4]. The effectiveness of colonoscopy largely depends on the quality of bowel preparation, with adequate cleanliness being essential for a safe, efficient examination and playing a critical role in clinical outcomes [6-8]. However, studies have reported that 20%‐40% of patients undergoing colonoscopy exhibit inadequate bowel preparation, which falls far below the European Society of Gastrointestinal Endoscopy–recommended adequacy standard of 90% [9-11]. Suboptimal bowel preparation may increase the risk of missed lesions, postprocedural complications, and prolonged procedural duration, and

XSL•FO
RenderX

necessitate repeat examinations, leading to additional patient discomfort and financial burden [12,13]. Thus, accurate identification of factors associated with bowel preparation holds significant clinical value for improving preparation quality. Numerous studies have identified multiple influencing factors that influence the quality of bowel preparation, including age, gender, BMI, anxiety level, and patient compliance with preparation protocols [14-20]. Strict compliance with bowel preparation instructions is recognized as a critical determinant of successful preparation. Poor compliance with prescribed requirements, including dietary restrictions, fluid intake limitations, and proper medication timing and dosage, may directly affect cleansing efficacy [19,20]. Current research predominantly evaluated compliance through single-dimensional metrics such as laxative consumption or dietary compliance, failing to holistically evaluate the full-process compliance, which includes dietary management, fluid intake restrictions, medication dosage, timing, and duration of medication use [21,22]. This study aims to evaluate the holistic compliance with bowel preparation instructions among patients undergoing colonoscopy and its multidimensional influencing factors in China. The findings will inform the optimization of personalized patient education strategies and the development of targeted preparation guidance, ultimately enhancing compliance to improve bowel preparation quality and subsequent colonoscopy outcomes.

Methods

Study Design

This cross-sectional study was conducted at West China Hospital from September 1, 2024, to March 1, 2025. Convenience sampling was used in this study.

Participants

The participants of this study were patients who underwent colonoscopy at the Endoscopy Center of West China Hospital, Sichuan University. At the time of scheduling a colonoscopy, standardized bowel preparation instructions were given verbally by the endoscopy unit receptionist, who was trained to provide these instructions as a daily routine. Patients and their relatives were also given a brochure as a reminder of bowel preparation instructions. The day before the colonoscopy, the patient began bowel preparation as instructed, including a low-fiber diet and the administration of a laxative, which was sodium phosphate or polyethylene glycol. Before undergoing the colonoscopy, patients waiting in the endoscopy center were invited to voluntarily and independently complete a mobile-based questionnaire by scanning a QR code. The questionnaire contained 32 single-choice items and required approximately 3 - 5 minutes to complete.

The following inclusion and exclusion criteria were used:

- Inclusion criteria: age ≥18 years, adequate understanding and communication abilities, and willingness to participate in the study and to sign the informed consent form
- Exclusion criteria: inability to understand the content of the informed consent form and the questionnaire, and failure to cooperate with the investigation or provide informed consent

Sample Size

The sample size was calculated using a single population proportion sample size–estimating formula. The formula can be given as $n = [(Z_{1-\alpha/2})^2 p (1-p)]/d^2$, where, n is the minimum sample size, $Z_{1-\alpha/2}$ is at a 99% CI of 2.58, *P* is the assumed adequate compliance with bowel preparation (50%), and d is the margin of error to be tolerated (5%). Using this formula, the estimated sample size is 666. Accounting for a 10% nonresponse rate, a total of 740 patients were needed. It should be noted that the 50% expected proportion was chosen as it provides the largest sample size estimate [23].

Questionnaire

General Information Questionnaire

The questionnaire was designed on the basis of a review of the literature and consultation with experts. It includes information on the following aspects: gender, age, educational level, residence, residency status, hypertension history, diabetes history, family history of CRC or polyps, know anyone with colorectal polyps or CRC other than a first-degree relative, and colonoscopy history.

Knowledge of Colonoscopy and Bowel Preparation

Patients' knowledge of colonoscopy and bowel preparation was measured using three and four questions, respectively. Each question has four response options: 0=no understanding at all, 1=a little understanding, 2=moderate understanding, and 3=complete understanding. Higher scores reflect better knowledge, with a score of 5 or above indicating adequate knowledge of colonoscopy and a score of 7 or above indicating adequate knowledge of bowel preparation.

Bowel Preparation Compliance

We designed a scale to assess patients' compliance with various aspects of bowel preparation instructions. The self-developed scale consists of 8 questions and is scored on a 4-point Likert scale, ranging from "completely noncompliant" to "completely compliant," with scores ranging from 0 to 3. The total score ranges from 0 to 24 points, with higher scores indicating greater bowel preparation compliance. A score greater than 95% of the total score was defined as adequate compliance and vice versa as inadequate compliance. Prior to the start of the formal study, we validated the questionnaire in a preexperiment that included 50 individuals, with a calculated Cronbach α coefficient of 0.877.

Current Bowel Preparation Experience Survey

In this section, we investigated the laxative type used by the patients for current bowel preparation, overall satisfaction with the taste of the laxative, anxiety during the bowel preparation, physical discomfort during the bowel preparation, and clarity of the bowel preparation instructions, with each variable classified as binary.

XSL•FO
RenderX

Study Variable Definitions

Dependent Variable

The dependent variable was the bowel preparation compliance (adequate or inadequate).

Independent Variables

The independent variables were gender; age; educational level (elementary school or no education, secondary school, university or higher); residence (urban, rural); residency status (living alone or with family); hypertension history; diabetes history; family history of CRC or polyps; know anyone with colorectal polyps or CRC other than a first-degree relative; colonoscopy history; knowledge of colonoscopy (adequate or inadequate); knowledge of bowel preparation (adequate or inadequate); and current bowel preparation experience, including overall satisfaction with the taste of the laxative (satisfactory or unsatisfactory), anxiety during the bowel preparation, physical discomfort during the bowel preparation, and clarity of the bowel preparation instructions.

Statistical Analysis

The data were analyzed using R version 4.3.2 (R Foundation for Statistical Computing). Continuous variables were presented as means and SDs, whereas categorical variables were expressed as numbers and percentages or frequencies. Intergroup differences with respect to continuous variables were assessed via Student *t* test, if the data met the normality assumption and variance homogeneity. Otherwise, the nonparametric Mann-Whitney *U* test was used. Intergroup differences with respect to categorical variables were assessed using Pearson χ^2 test or Fisher exact probability test, when appropriate.

Variables with $P<.10$ in the univariate analysis were included in the multivariate logistic regression analysis, and independent factors influencing patients' compliance with bowel preparation were identified through the "enter" selection method (ie, the totally adjusted model). The Hosmer-Lemeshow goodness-of-fit test was used to assess the fitness of the model, with $P>.05$ indicating an acceptable fit. The multivariate analysis results are expressed using odds ratios and 95% CIs only. A $P<.05$ was regarded as statistically significant.

Ethical Considerations

The research study received approval from the Ethics Committee of West China Hospital, Sichuan University (ChiCTR2400089364). Prior to conducting the questionnaire, written consent was obtained from all participants. The privacy of the participants was respected. All data in the manuscript were anonymized in accordance with ethical standards, ensuring no personally identifiable information could be discerned. No compensation was provided to participants in this study.

Results

Sample Characteristics

This study ultimately included 740 patients undergoing colonoscopy. Among the total sample, more than half (n=416, 56.2%) were female, 70% (n=518) had a university education or higher, and 90.5% (n=670) lived in urban areas. The majority (n=704, 95.1%) had no personal history of diabetes, and 86.5% (n=640) had no family history of CRC or polyps. More than half (n=385, 52.0%) of the patients had no prior colonoscopy. Other variables are listed in Table 1.

XSL•FO
RenderX

Table . General information of patients undergoing colonoscopy (N=740).

Variables	Patients, n (%)
Gender	
Male	324 (43.8)
Female	416 (56.2)
Age (years)	
18-35	152 (20.5)
36 - 49	245 (33.1)
50 - 64	260 (35.1)
65 - 74	75 (10.1)
≥75	8 (1.1)
Educational level	
Elementary school or no education	35 (4.7)
Secondary school	187 (25.3)
University or higher	518 (70.0)
Residence	
Urban	670 (90.5)
Rural	70 (9.5)
Residency status	
Living alone	45 (6.1)
Living with family	695 (93.9)
Diabetes history	
No	704 (95.1)
Yes	36 (4.9)
Hypertension history	
No	632 (85.4)
Yes	108 (14.6)
Family history of colorectal cancer or polyps	
No	640 (86.5)
Yes	100 (13.5)
Know anyone with colorectal polyps or colorectal cancer other than a first-degree relative	
No	444 (60.0)
Yes	296 (40.0)
Colonoscopy history	
No	385 (52.0)
Yes	355 (48.0)

Knowledge of Bowel Preparation and Colonoscopy and Current Bowel Preparation Experience of Patients Undergoing Colonoscopy

Table 2 presents colonoscopy patients' knowledge of bowel preparation and colonoscopy, as well as their experience with the current bowel preparation. Of the 740 patients, a total of 56.4% (n=417) and 46.1% (n=341) had adequate knowledge of bowel preparation and colonoscopy, respectively. More than half (n=391, 52.8%) of the patients were unsatisfied with the taste of the laxative. Anxiety and physical discomfort during bowel preparation were experienced by 43.9% (n=325) and 59.3% (n=439) of patients, respectively.

XSL•FO
RenderX

Table . Knowledge of bowel preparation and colonoscopy and current bowel preparation experience of patients undergoing colonoscopy (N=740).

Variables	Patients, n (%)
Knowledge of bowel preparation	
Inadequate or poor	323 (43.6)
Adequate or excellent	417 (56.4)
Knowledge of colonoscopy	
Inadequate or poor	399 (53.9)
Adequate or excellent	341 (46.1)
Any bowel discomfort prior to this colonoscopy	
No	350 (47.3)
Yes	390 (52.7)
Current colonoscopy schedule	
Morning	244 (33.0)
Afternoon	309 (41.8)
Evening	187 (25.3)
Laxative type	
Sodium phosphate	246 (33.2)
Polyethylene glycol	494 (66.8)
Satisfaction with the taste of the laxative	
Unsatisfactory	391 (52.8)
Satisfactory	349 (47.2)
Anxiety during bowel preparation	
No	415 (56.1)
Yes	325 (43.9)
Physical discomfort during bowel preparation	
No	301 (40.7)
Yes	439 (59.3)
Clarity of bowel preparation instructions	
No	19 (2.6)
Yes	721 (97.4)

Univariate Analysis of Factors Influencing Compliance With Bowel Preparation

Of the 740 patients, 311 (42.0%) showed adequate compliance with bowel preparation. Intergroup differences were compared between patients with and without adequate bowel preparation compliance (Table 3). The results indicate that hypertension history (P=.01), knowledge of bowel preparation (P=.004), laxative type (P=.008), satisfaction with the taste and flavor of the laxative (P<.001), anxiety during bowel preparation (P<.001), and physical discomfort during bowel preparation (P<.001) all have statistically significant impacts on patient compliance (Table 3).

XSL•FO
RenderX

Table . Univariate analysis of bowel preparation compliance in patients undergoing colonoscopy.

Variables	Adequate compliance (n=311)	Inadequate compliance (n=429)	Chi-square (*df*)	*P* value
Gender			0.04 (1)	.84
Male	138 (44.4)	186 (43.4)		
Female	173 (55.6)	243 (56.6)		
Age (years)			9.22 (4)	.06
18 - 35	57 (18.3)	95 (22.1)		
36 - 49	94 (30.2)	151 (35.2)		
50 - 64	115 (37.0)	145 (33.8)		
65 - 74	42 (13.5)	33 (7.7)		
≥75	3 (1.0)	5 (1.2)		
Educational level			5.97 (2)	.05
Elementary school or no education	9 (2.9)	26 (6.1)		
Secondary school	72 (23.2)	115 (26.8)		
University or higher	230 (74.0)	288 (67.1)		
Residence			1.00 (1)	.32
Urban	286 (92.0)	384 (89.5)		
Rural	25 (8.0)	45 (10.5)		
Residency status			0.19 (1)	.66
Living alone	17 (5.5)	28 (6.5)		
Living with family	294 (94.5)	401 (93.5)		
Diabetes history			0.02 (1)	.90
No	295 (94.9)	409 (95.3)		
Yes	16 (5.1)	20 (4.7)		
Hypertension history			6.53 (1)	.01[a]
No	253 (81.4)	379 (88.3)		
Yes	58 (18.6)	50 (11.7)		
Family history of colorectal cancer or polyps			0.11 (1)	.74
No	271 (87.1)	369 (86.0)		
Yes	40 (12.9)	60 (14.0)		
Know anyone with colorectal polyps or colorectal cancer other than a first-degree relative			1.52 (1)	.22
No	178 (57.2)	266 (62.0)		
Yes	133 (42.8)	163 (38.0)		
Colonoscopy history			0.04 (1)	.85
No	160 (51.4)	225 (52.4)		
Yes	151 (48.6)	204 (47.6)		
Knowledge of bowel preparation			8.35 (1)	.004[a]
Inadequate or poor	116 (37.3)	207 (48.3)		
Adequate or excellent	195 (62.7)	222 (51.7)		
Knowledge of colonoscopy			1.88 (1)	.17
Inadequate or poor	158 (50.8)	241 (56.2)		
Adequate or excellent	153 (49.2)	188 (43.8)		
Any bowel discomfort prior to this colonoscopy			0.00 (1)	>.99

XSL•FO
RenderX

Variables	Adequate compliance (n=311)	Inadequate compliance (n=429)	Chi-square (*df*)	*P* value
No	147 (47.3)	203 (47.3)		
Yes	164 (52.7)	226 (52.7)		
Current colonoscopy schedule			4.01 (2)	.14
Morning	90 (28.9)	154 (35.9)		
Afternoon	139 (44.7)	170 (39.6)		
Evening	82 (26.4)	105 (24.5)		
Laxative type			7.13 (1)	.008[a]
Sodium phosphate	86 (27.7)	160 (37.3)		
Polyethylene glycol	225 (72.3)	269 (62.7)		
Satisfaction with the taste of the laxative			50.90 (1)	<.001[a]
Unsatisfactory	116 (37.3)	275 (64.1)		
Satisfactory	195 (62.7)	154 (35.9)		
Anxiety during bowel preparation			21.80 (1)	<.001[a]
No	206 (66.2)	209 (48.7)		
Yes	105 (33.8)	220 (51.3)		
Physical discomfort during bowel preparation			64.60 (1)	<.001[a]
No	180 (57.9)	121 (28.2)		
Yes	131 (42.1)	308 (71.8)		
Clarity of bowel preparation instructions			2.69 (1)	.10
No	4 (1.3)	15 (3.5)		
Yes	307 (98.7)	414 (96.5)		

[a]*P*<.05.

Multivariate Analysis of Factors Influencing Bowel Preparation Compliance

Binary logistic regression analysis indicated that age, educational level, knowledge of bowel preparation, satisfaction with the taste of the laxative, and physical discomfort during bowel preparation are independent influencing factors for patient compliance (Table 4). The Hosmer-Lemeshow goodness-of-fit test was not significant, indicating that the model fits the data well (χ^2_8=5.09; *P*=.75).

XSL•FO
RenderX

Table . Multivariate analysis of bowel preparation compliance in patients undergoing colonoscopy.

Variables	Odds ratio (95% CI)	*P* value
Age (years)		
18 - 35	Reference	—[a]
36 - 49	1.04 (0.67-1.64)	.85
50 - 64	1.23 (0.76-1.98)	.41
65 - 74	2.27 (1.16-4.49)	.02[b]
≥75	0.91 (0.16-4.50)	.91
Educational level		
Elementary school or no education	Reference	—
Secondary school	1.99 (0.85-5.00)	.12
University or higher	3.29 (1.41-8.33)	.008[b]
Hypertension history		
No	Reference	—
Yes	1.56 (0.95-2.56)	.08
Knowledge of bowel preparation		
Inadequate or poor	Reference	—
Adequate or excellent	1.59 (1.14-2.24)	.007[b]
Laxative type		
Sodium phosphate	Reference	—
Polyethylene glycol	1.15 (0.81-1.63)	.44
Satisfaction with the taste and flavor of the laxative		
Unsatisfactory	Reference	—
Satisfactory	2.11 (1.48-3.02)	<.001[b]
Anxiety during bowel preparation		
No	Reference	—
Yes	0.82 (0.57-1.18)	.28
Physical discomfort during bowel preparation		
No	Reference	—
Yes	0.45 (0.31-0.64)	<.001[b]

[a]Not applicable.
[b]*P*<.05.

Discussion

One of our study objectives is to evaluate the compliance rate with bowel preparation of Chinese patients undergoing colonoscopy. The survey results revealed that inadequate bowel preparation compliance among colonoscopy patients was up to 58.0% (429/740), which was significantly higher than the 23.6% previously reported in a study [24]. This may be related to the more multidimensional (including dietary management, fluid intake restrictions, medication dosage, timing, and duration of medication use) and demanding (a score greater than 95% of the total compliance score) assessment of bowel preparation compliance in our study.

Our results showed that patients' compliance with bowel preparation instructions was related to age, educational level, and knowledge of bowel preparation. Patients in the 65 - 74 years age group had higher compliance than younger patients in the 18 - 35 years age group. This may be related to the fact that the retired population has more time to read bowel preparation instructions, whereas the younger patients may ignore relevant details due to work pressure and busy routines. Moreover, the busy lifestyle and tendency of young people to eat out may also hinder their strict compliance with the long and complex food list for dietary restriction before colonoscopy [25]. Patients with a higher level of education and better knowledge of bowel preparation are more likely to understand the educational content provided by health care professionals

XSL•FO
RenderX

[26,27]. These patients may tend to recognize the importance and benefits of adequate bowel preparation for colonoscopy, thus improving their compliance [24]. In the future, personalized health education content can be developed for patients on the basis of their age and educational level to increase their knowledge and improve compliance. For example, customized instructions with concise content can be designed specifically for young people to address their unique challenges.

In this study, the type of laxative significantly influenced compliance in univariate analysis. However, in multivariate analysis, it no longer emerged as an independent factor, suggesting that the effect of laxative type may primarily influence compliance indirectly through associated experiential factors such as taste satisfaction and physical discomfort. Sodium phosphate and polyethylene glycol are two commonly used clinical laxatives that commonly cause adverse reactions such as bloating, nausea, vomiting, and dizziness [28]. These adverse reactions, as well as poor taste and flavor, may result in patients not being able to take adequate amounts of laxative, leading to poor compliance. Of the 740 patients in this study, more than half (n=391, 52.8%) were unsatisfied with the taste of the laxative used, and close to 60% (n=439) experienced physical discomfort during bowel preparation. A study by Hautefeuille et al [29] also showed that patients described nausea and vomiting as the main reasons for noncompliance with bowel preparation protocols. Therefore, for patients with prior experience of bowel preparation, it is important to inquire about their previous experiences with laxatives when they need to undergo bowel preparation again, and laxatives that are acceptable to the patient in terms of taste should be selected to minimize the occurrence of adverse reactions and improve patient compliance.

There are several limitations of this study. First, this study adopted a convenience sampling method and was conducted in a single tertiary hospital in an urban area, which may have introduced selection bias and led to an overestimation of certain indicators (eg, educational level and health awareness of patients). Future research could expand the study to multiple centers to gain a more comprehensive understanding of patient compliance in different settings. Second, bowel preparation compliance was self-reported, which may introduce recall bias, potentially leading to overestimation or underestimation of actual compliance. In addition, it was assessed using a newly developed scale that has only undergone preliminary validation, and its reliability and generalizability require further confirmation in larger and more diverse populations. Third, this study did not collect results on the quality of patients' bowel preparation, nor did it explore the correlation between compliance and bowel preparation quality. Lastly, due to the cross-sectional design, it is not possible to establish causal relationships. Overall, these limitations highlight critical areas for future research and improvement in understanding and bowel preparation compliance in patients undergoing colonoscopy.

This study revealed that patients had relatively low compliance with bowel preparation. Independent factors identified included age, educational level, knowledge of bowel preparation, satisfaction with the taste of the laxative, and physical discomfort during bowel preparation. Based on the data from this study, we recommend personalizing the educational content of bowel preparation according to the patients' age and literacy level, and selecting the type of laxative that the patient is satisfied with when available. Our study provides insights for improving bowel preparation compliance in patients undergoing colonoscopy.

Acknowledgments

This study was supported by the Science and Technology Project of Sichuan Provincial Health Commission (grant 23LCYJ037) and Sichuan Province Cadres Health Research (grant ZH2024 -102).

Data Availability

The data presented in this study are available upon request from the corresponding author. The data are not publicly available due to privacy restrictions.

Authors' Contributions

HJ and BH conceived and designed the study. HJ and CL carried out the data collection. HJ analyzed the data. HJ drafted the manuscript. BH and YM reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Feng L, Guan J, Dong R, et al. Risk factors for inadequate bowel preparation before colonoscopy: a meta-analysis. J Evid Based Med 2024 Jun;17(2):341-350. [doi: 10.1111/jebm.12607] [Medline: 38651546]
2. Sulz MC, Kröger A, Prakash M, Manser CN, Heinrich H, Misselwitz B. Meta-analysis of the effect of bowel preparation on adenoma detection: early adenomas affected stronger than advanced adenomas. PLoS ONE 2016;11(6):e0154149. [doi: 10.1371/journal.pone.0154149] [Medline: 27257916]

XSL•FO
RenderX

3. AlAmeel T. Bowel preparation: the elderly, the hospitalized, and the colonoscope. Saudi J Gastroenterol 2018;24(2):73-74. [doi: 10.4103/sjg.SJG_137_18] [Medline: 29637912]
4. Haghighat S, Sussman DA, Deshpande A. US Preventive Services Task Force recommendation statement on screening for colorectal cancer. JAMA 2021 Oct 5;326(13):1328. [doi: 10.1001/jama.2021.13466] [Medline: 34609459]
5. Gravina AG, Pellegrino R, Romeo M, et al. Quality of bowel preparation in patients with inflammatory bowel disease undergoing colonoscopy: what factors to consider? World J Gastrointest Endosc 2023 Mar 16;15(3):133-145. [doi: 10.4253/wjge.v15.i3.133] [Medline: 37034970]
6. Sharma VK, Steinberg EN, Vasudeva R, Howden CW. Randomized, controlled study of pretreatment with magnesium citrate on the quality of colonoscopy preparation with polyethylene glycol electrolyte lavage solution. Gastrointest Endosc 1997 Dec;46(6):541-543. [doi: 10.1016/s0016-5107(97)70011-7] [Medline: 9434223]
7. van Doorn SC, Dekker E. Colonoscopy quality begins with a clean colon. Endoscopy 2012 Jul;44(7):639-640. [doi: 10.1055/s-0032-1309975] [Medline: 22723181]
8. Lee JK, Jensen CD, Levin TR, et al. Long-term risk of colorectal cancer and related death after adenoma removal in a large, community-based population. Gastroenterology 2020 Mar;158(4):884-894. [doi: 10.1053/j.gastro.2019.09.039] [Medline: 31589872]
9. Jawa H, Mosli M, Alsamadani W, et al. Predictors of inadequate bowel preparation for inpatient colonoscopy. Turk J Gastroenterol 2017 Nov;28(6):460-464. [doi: 10.5152/tjg.2017.17196] [Medline: 29086713]
10. Serper M, Gawron AJ, Smith SG, et al. Patient factors that affect quality of colonoscopy preparation. Clin Gastroenterol Hepatol 2014 Mar;12(3):451-457. [doi: 10.1016/j.cgh.2013.07.036] [Medline: 23954647]
11. Hassan C, Bretthauer M, Kaminski MF, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) guideline. Endoscopy 2013;45(2):142-150. [doi: 10.1055/s-0032-1326186] [Medline: 23335011]
12. Byrne MF. The curse of poor bowel preparation for colonoscopy. Am J Gastroenterol 2002 Jul;97(7):1587-1590. [doi: 10.1111/j.1572-0241.2002.05855.x] [Medline: 12135005]
13. Johnson DA, Barkun AN, Cohen LB, et al. Optimizing adequacy of bowel cleansing for colonoscopy: recommendations from the US Multi-Society Task Force on Colorectal Cancer. Am J Gastroenterol 2014 Oct;109(10):1528-1545. [doi: 10.1038/ajg.2014.272] [Medline: 25223578]
14. Wu KL, Rayner CK, Chuah SK, Chiu KW, Lu CC, Chiu YC. Impact of low-residue diet on bowel preparation for colonoscopy. Dis Colon Rectum 2011 Jan;54(1):107-112. [doi: 10.1007/DCR.0b013e3181fb1e52] [Medline: 21160321]
15. Wong MCS, Ching JYL, Chan VCW, et al. Determinants of bowel preparation quality and its association with adenoma detection: a prospective colonoscopy study. Medicine (Baltimore) 2016 Jan;95(2):e2251. [doi: 10.1097/MD.0000000000002251] [Medline: 26765402]
16. Borg BB, Gupta NK, Zuckerman GR, Banerjee B, Gyawali CP. Impact of obesity on bowel preparation for colonoscopy. Clin Gastroenterol Hepatol 2009 Jun;7(6):670-675. [doi: 10.1016/j.cgh.2009.02.014] [Medline: 19245852]
17. Kutlutürkan S, Görgülü U, Fesci H, Karavelioglu A. The effects of providing pre-gastrointestinal endoscopy written educational material on patients' anxiety: a randomised controlled trial. Int J Nurs Stud 2010 Sep;47(9):1066-1073. [doi: 10.1016/j.ijnurstu.2010.01.007] [Medline: 20181334]
18. Amitay EL, Niedermaier T, Gies A, Hoffmeister M, Brenner H. Risk factors of inadequate bowel preparation for screening colonoscopy. J Clin Med 2021 Jun 21;10(12):2740. [doi: 10.3390/jcm10122740] [Medline: 34205800]
19. Seo EH, Kim TO, Park MJ, et al. Optimal preparation-to-colonoscopy interval in split-dose PEG bowel preparation determines satisfactory bowel preparation quality: an observational prospective study. Gastrointest Endosc 2012 Mar;75(3):583-590. [doi: 10.1016/j.gie.2011.09.029] [Medline: 22177570]
20. Fang J, Fu HY, Ma D, et al. Constipation, fiber intake and non-compliance contribute to inadequate colonoscopy bowel preparation: a prospective cohort study. J Dig Dis 2016 Jul;17(7):458-463. [doi: 10.1111/1751-2980.12376] [Medline: 27356275]
21. Liu X, Luo H, Zhang L, et al. Telephone-based re-education on the day before colonoscopy improves the quality of bowel preparation and the polyp detection rate: a prospective, colonoscopist-blinded, randomised, controlled study. Gut 2014 Jan;63(1):125-130. [doi: 10.1136/gutjnl-2012-304292] [Medline: 23503044]
22. Lee YJ, Kim ES, Choi JH, et al. Impact of reinforced education by telephone and short message service on the quality of bowel preparation: a randomized controlled study. Endoscopy 2015 Nov;47(11):1018-1027. [doi: 10.1055/s-0034-1392406] [Medline: 26182387]
23. ALruwaili BF. Evaluation of hypertension-related knowledge, medication adherence, and associated factors among hypertensive patients in the Aljouf Region, Saudi Arabia: a cross-sectional study. Medicina (Kaunas) 2024 Nov 6;60(11):1822. [doi: 10.3390/medicina60111822] [Medline: 39597007]
24. Chan WK, Saravanan A, Manikam J, Goh KL, Mahadeva S. Appointment waiting times and education level influence the quality of bowel preparation in adult patients undergoing colonoscopy. BMC Gastroenterol 2011 Jul 28;11:86. [doi: 10.1186/1471-230X-11-86] [Medline: 21798022]
25. Chou CK, Chang CY, Chang CC, et al. Controlled dietary restriction with a prepackaged low-residue diet before colonoscopy offers better-quality bowel cleansing and allows the use of a smaller volume of purgatives: a randomized multicenter trial. Dis Colon Rectum 2016 Oct;59(10):975-983. [doi: 10.1097/DCR.0000000000000675] [Medline: 27602929]

XSL•FO
RenderX

26. Nguyen DL, Wieland M. Risk factors predictive of poor quality preparation during average risk colonoscopy screening: the importance of health literacy. J Gastrointestin Liver Dis 2010 Dec;19(4):369-372. [Medline: 21188326]
27. Smith SG, von Wagner C, McGregor LM, et al. The influence of health literacy on comprehension of a colonoscopy preparation information leaflet. Dis Colon Rectum 2012 Oct;55(10):1074-1080. [doi: 10.1097/DCR.0b013e31826359ac] [Medline: 22965407]
28. Unger RZ, Amstutz SP, Seo DH, Huffman M, Rex DK. Willingness to undergo split-dose bowel preparation for colonoscopy and compliance with split-dose instructions. Dig Dis Sci 2010 Jul;55(7):2030-2034. [doi: 10.1007/s10620-009-1092-x] [Medline: 20082216]
29. Hautefeuille G, Lapuelle J, Chaussade S, et al. Factors related to bowel cleansing failure before colonoscopy: results of the PACOME study. United European Gastroenterol J 2014 Feb;2(1):22-29. [doi: 10.1177/2050640613518200] [Medline: 24918005]

Abbreviations

CRC: colorectal cancer
OR: odds ratio

Edited by N Cahill, TDA Cardoso; submitted 08.05.25; peer-reviewed by J Soldera, N Rattanapitoon, P Shingru; revised version received 29.09.25; accepted 30.09.25; published 21.10.25.

Please cite as:
Jiang H, Li C, Hu B, Mou Y
Compliance With Bowel Preparation and Its Influencing Factors in Patients Undergoing Colonoscopy: Cross-Sectional Study
Interact J Med Res 2025;14:e77189
URL: https://www.i-jmr.org/2025/1/e77189
doi:10.2196/77189

© Huan Jiang, Chuanhui Li, Bing Hu, Yi Mou. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 21.10.2025. 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 Color of Drinking Survey Questionnaire for Measuring the Secondhand Impacts of High-Risk Drinking in College Settings: Validation Study

Agustina Marconi[1*], MPH, MD; Reonda Washington[1*], MPH; Amanda Jovaag[1*], MS; Courtney Blomme[1*], MS, RD; Ashley Knobeloch[1*], BNS, MPH, RN; Vilma Irazola[2*], MSc, MD; Carolina Muros Cortés[2*], MD; Laura Gutierrez[2*], MSc; Natalia Elorriaga[2*], LN, MSc, PhD

[1]UW-Madison, 333 East Campus Mall 8th floor, Madison, WI, United States

[2]Instituto de Efectividad Clinica y Sanitaria (IECS), Dr. Emilio Ravignani 2024, Buenos Aires, Argentina

[*]all authors contributed equally

Corresponding Author:
Agustina Marconi, MPH, MD
UW-Madison, 333 East Campus Mall 8th floor, Madison, WI, United States

Abstract

Background: The "Color of Drinking" is a study conducted at the University of Wisconsin-Madison. It examines the secondhand harms of high-risk drinking on college students of color and explores the connection between alcohol use and the campus racial climate. Since its findings were released in 2018, this study has received significant attention from other college settings around the country.

Objective: This study aims to describe the development of the most recent version of the Color of Drinking questionnaire and to assess its internal consistency, test-retest reliability, and construct validity in a sample of undergraduate students attending the University of Wisconsin-Madison.

Methods: This is an observational, analytic study. Questionnaire design experts revised the original instrument, and in-depth cognitive interviews with students were conducted to evaluate comprehensibility and acceptability. The revised questionnaire was administered 2 times, 3 to 4 weeks apart, in a sample of undergraduate students. The following properties were studied: internal consistency in 4 sets of items (Cronbach α), test-retest reliability among closed-ended questions (κ statistics and intraclass correlation coefficient), and construct validity (associations with other validated instruments, such as the Alcohol Use Disorders Identification Test). For a section of questions showing low reliability, the answers to open questions and other in-depth interviews were carried out, and online surveys were conducted with another sample of undergraduate students to evaluate reliability after changes.

Results: Eight students participated in the in-depth interviews, 177 responses from the online survey were included for the analysis of internal consistency, 115 for test-retest reliability, and 98 for construct validity. The 4 sets of items (sections) evaluated ("impact of alcohol consumption on academics," "impact of microaggressions," "witnessing microaggressions and alcohol intoxication," and "bystanders' interventions on alcohol intoxication") presented good internal consistency (Cronbach α between 0.723 and 0.898). Most items showed moderate to substantial test-retest reliability; agreement was from 68.1% to 95.2%, and κ coefficients ranged from 0.214 to 0.8. For construct validity, correlations between the number of drinking days, the maximum number of drinks in a day and the Alcohol Use Disorders Identification Test score were moderate to high, r=0.630 (95% CI 0.533-0.719) and r=0.647 (95% CI 0.548-0.741), respectively. Due to low reliability, a section regarding "health impacts" has been redesigned, including 8 items for the personal consumption of alcohol and the consumption of others (Cronbach α 0.735 and 0.855, respectively; agreement between the first and the second time the questionnaire was administered were 83.4% and 99.1%, and most of the items with κ coefficient from 0.476 to 0.877).

Conclusions: The revised version of the Color of Drinking questionnaire showed acceptable to adequate reliability and construct validity.

(Interact J Med Res 2025;14:e64720) doi:10.2196/64720

KEYWORDS

validation study; alcohol drinking in college; microaggression; university; student; young adult; undergraduate; survey; questionnaire; reliability; consistency

XSL•FO
RenderX

Introduction

The association between alcohol consumption and being a college student is broadly described in the literature. Students often understand drinking during college as part of their higher education experience and a way to socialize with peers [1]. The first years of college involve major transitions, including changes in family bonds, living arrangements, and peer socialization. That type of transition is often associated with risky behaviors, including high levels of alcohol consumption [2,3]. Environmental factors, such as the campus culture, can further [2,3] reinforce this behavior and contribute to its health consequences [4]. However, few studies have examined the potential impact of a high-risk drinking culture on the mental health and well-being of marginalized college students [5] The Color of Drinking study examines differences in narratives and perspectives related to alcohol culture, high-risk drinking, and the racial climate for students of color and White students at an American predominately White institution [6].

Results from the exploratory Color of Drinking Study (2017/2018) revealed key insights into the intersection of alcohol culture and student experiences. Safety concerns in high-risk drinking environments had a greater impact on students of color compared to their White peers. High-risk drinking was also positively associated with students' sense of connection and belonging. Students of color reported higher rates of abstaining from or avoiding alcohol than White students (15.3% vs 8.4%). Additionally, African American or Black students considered leaving the university at twice the rate of White students (43.6% vs 20.9%), and students of color were more likely to report financial struggles (14.5% vs 7%). White, problematic drinkers scored higher on the Diener Flourishing Scale than students of color (27.3% vs 10.2%). Campus alcohol culture also influenced academic routines, with 46% of students reporting the need to find alternative study spaces and 40% scheduling group meetings around alcohol consumption [6]. Furthermore, African American or Black was the racial group that most frequently witnessed (82.9%) and experienced microaggressions (79.8%). The likelihood of witnessing microaggressions increased with the year in school, from 49.3% among first-year students to 75.8% among fourth-year students. Among students of color, alcohol use in the last 30 days, feeling impacted by other's consumption of alcohol, and avoiding certain areas due to alcohol consumption were all significantly associated with experiencing microaggressions among students of color [7].

Since the release of its findings, the Color of Drinking has received significant attention from other college settings, media coverage, and numerous requests to replicate this study around the country [8]. As the instrument gained greater prominence as a research tool and a way for campuses to understand their student experiences, the decision was made to conduct a validation study of the Color of Drinking instrument. This validation effort enhances the University of Wisconsin-Madison's (UW-Madison's) ability to interpret survey results accurately and provides other institutions with a clearer understanding of the insights they can gain from administering the survey.

Thus, this study aims to describe the development of the last version of the Color of Drinking questionnaire and to evaluate its reliability and validity in a sample of undergraduate students attending UW-Madison.

Methods

Study Design

This observational, analytic study incorporated both qualitative and quantitative research components. Experts in questionnaire design reviewed the original version of the questionnaire and in-depth cognitive interviews were conducted with students to evaluate the comprehensibility and acceptability of the instrument. The revised version of the questionnaire was administered to a sample of undergraduate students at UW-Madison twice, with an interval of 3 to 4 weeks. This timeframe was selected to reduce the likelihood of participants remembering their previous responses while avoiding a gap long enough to span different semesters, which could impact the comparability of recall. Other validated instruments were administered alongside the first administration of the Color of Drinking questionnaire: (1) the Alcohol Use Disorders Identification Test (AUDIT), the microaggressions subscale from the Racial and Ethnic Microaggression Scale (REMS) [9], and the Sense of Belonging Index (SBI). The following properties were evaluated: internal consistency, test-retest reliability, and construct validity.

Participants

The study was conducted in a sample of undergraduate students aged ≥18 years at UW-Madison.For in-depth cognitive interviews, an opportunistic sample of students (n=8), ensuring variability in age, gender, and ethnicity or race among students with freshman, sophomore, junior, and senior standing, was selected. For the online survey the minimum sample size was estimated in 100 participants. This sample size was determined to assess the construct validity by applying a 2-sample *t* test, based on data previously published from the REMS subscale in a sample of undergraduate students, considering a mean difference of 0.5, an SD of 0.89, an error α of .05 and a statistical power (1-β) of .80 [10]. This sample size is also adequate for assessing test-retest reliability (n=63 participants completing the survey on 2 occasions), based on an anticipated intraclass correlation of 0.7 (with a minimum acceptably threshold of 0.5 [11]), a significance level (α) of .05 and β of .20 [12,13].

Student identity data was obtained from the UW-Madison Office of the Registrar and sampled self-identified White and censused self-identified undergraduate students of color or international students. Based on the recruitment rate from the study previously conducted in the same population [7], a random sample of 300 self-identified White undergraduate students and 200 self-identified undergraduate students of color or international students were invited to participate. Each student received an email with the invitation to participate, and 2 reminder emails were sent. It was explained that there were 2 opportunities to be part of the validation of the Color of Drinking Survey (a qualitative Zoom [Zoom Communications, Inc] interview or an online survey administered twice). Participation was voluntary,

XSL•FO
RenderX

and invitations were sent via listservs. Thus, students had the option not to respond to the task. Participants who had completed the first questionnaire were contacted by email with the invitation to complete the second questionnaire.

Color of Drinking Questionnaire

The original instrument was developed by the University Health Services at the UW-Madison [6]. The dimensions of interest were selected, and a list of questions was developed by an interdisciplinary team composed of a Subject Matter Expert on Alcohol (RW) and a Physician, Epidemiologist specialist in Public Health (AM), based on information gathered during interviews with students and a framework based on previous research [5]. The instrument explored the following dimensions: (1) personal alcohol consumption, (2) academics and alcohol drinking culture, (3) locations (avoiding areas at campus due to concerns of the alcohol use of others or where feeling unsafe), (4) the impact of others' alcohol use, (5) experiencing microaggressions (targeting students of color only), (6) bystander witnessing and intervention in situations of microaggressions or alcohol intoxications), (7) impact on personal health and well-being, and (8) overall experience and sense of belonging at the campus. The tool also included questions about sociodemographic characteristics such as age, gender, race or ethnicity, and financial situation. The questionnaire included both open-ended and closed-ended questions. Questions about experiencing microaggressions were asked only to students of color, under the definition of microaggressions used by the authors: "racial microaggressions are an everyday manifestation of oppression that brings psychological consequences to target groups" [14]. The well-being section included a validated scale, the Diener Flourishing Scale, consisting of 8 items [15]. The questionnaire was previously administered in 2015, 2017, and 2018 to more than 1100 students; thus, the research team has previously analyzed useful information from both closed-ended and open-ended questions.

During the first phase of the validation study, the original version was revised by a group of researchers with expertise in questionnaire design, obtaining a preliminary draft version. Later, a series of cognitive interviews were conducted with a sample of students. They consisted of the flexible administration of the draft questionnaire and relied heavily on verbal probing to improve the instrument [16,17]. An "inspect-and-repair" model was used to inspect the items, detect flaws, and repair them. If the interviewer identified any problem with 1 or more questions (about comprehension, recall, difficulty to answer, response, or other), they would rephrase the items to test a potential solution [18]. The interviews were focused on the questions developed for the questionnaire and did not include the already validated Diener Flourishing Scale or the section on sociodemographic characteristics. After the series of interviews, a new version of the questionnaire was produced.

Assessment of the Psychometric Properties

Internal Consistency

This property is a measure of the intercorrelation among items and hence the consistency in the measurement of the intended construct [19]. The internal consistency was evaluated in the following sets of items: (1) impact of alcohol consumption on academics (Q8. How often have you experienced the following during the current semester? items: 8.1 to 8.5), (2) impact of microaggressions (Q17. How much did the microaggressions impact your… items: 17.1 to 17.3?), (3) witnessed microaggressions (Q25. Have you witnessed any of the following? ...items: 25.1 to 25.8), and (4) bystander intervention (Q26. Would you intervene in the following situations? ...items: 26.1 to 26.4)

Test-Retest Reliability

Test-retest reliability, a measurement of reproducibility, was evaluated for closed-ended questions.

Construct Validity

To assess construct validity, participants also completed 3 validated scales: the AUDIT [20], the SBI [21], and the "Workplace and School Microaggressions" Scale from the REMS [9]. They also reported their Grade Point Average (GPA).

The AUDIT is a 10-item questionnaire with a range of possible scores from 0 to 40. Scores ≥7 and 8 have been proposed as "hazardous and harmful alcohol use" among women and men, respectively [20,22]. It was expected that (1) the median number of drinking days and the maximum number of drinks in a day during the last 2 weeks measured by the Color of Drinking questionnaire are greater among those scoring ≥7 than those scoring <7 in the AUDIT instrument, and (2) the number of drinking days and the maximum number of drinks per day during the last 2 weeks are correlated with the AUDIT score.

The validated "Workplace and School Microaggressions" subscale from the REMS, included 5 items describing experienced microaggressions in the last 6 months, with possible score ranges from 0 to 5. It was expected that those who answered that they had experienced a microaggression on campus or surrounding areas (Color of Drinking, Q14) score higher in the REMS subscale than those who did not have such experience [9].

The SBI is a validated 6-item questionnaire with possible scores ranging from 5 to 30 (the greater the score, the more the sense of belonging). It was hypothesized that the SBI score is lower among those who reported that the alcohol culture had impacted their overall sense of belonging at UW-Madison in comparison to those who reported that the alcohol culture had not affected it [21].

Finally, the academic performance was operationalized as self-reported accumulated and last-term GPA. It was expected that academic performance was better among those students who responded that they never received a poor grade because they had chosen to drink instead of study during the semester.

Redesign and Assessment for Questions With Low Performance and Developing the Final Questionnaire

For a section of questions showing low reliability, the answers to the open-ended questions and other series of in-depth interviews were used to generate a revised set of questions.

XSL•FO
RenderX

Then, online surveys were conducted with another sample of undergraduate students to evaluate reliability after changes.

Data Analysis

Descriptive statistics were calculated for the population and each item. Hypothesis tests were selected according to the variable distribution. Wilcoxon rank sum test with continuity correction and Kendall correlation coefficients were used to test the hypotheses related to construct validity.

The internal consistency was evaluated by calculating the Cronbach α coefficient by dimension. Additionally, we calculated the variation in the α coefficient after excluding each item from the analysis to assess the effect of dropping a particular item. A Cronbach α value of at least 0.7 is usually interpreted as adequate internal consistency [23]. Low values of α represent poor interrelatedness between items.

To assess test-retest reliability, the Cohen κ coefficient (dichotomous data), the Cohen weighted κ coefficient (ordinal data), and the intraclass correlation coefficient (for continuous data) and their 95% CI were calculated. The higher the coefficient, the stronger the test-retest reliability [13]. κ values between 0.41 and 0.6 were considered to be of moderate reliability, between 0.61 and 0.8 as substantial reliability and between 0.81 and 1 very high reliability [19]. κ is influenced by the trait prevalence and basal rates. κ may be low even though there are high levels of agreement and even though individual ratings are accurate, thus we have also calculated the proportion of observed and expected agreement only by chance. For the interpretation of the intraclass correlation coefficient, values lower than 0.5 were considered indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 good reliability, and values higher than 0.9 excellent reliability [11].

All the analyses considered $\alpha < .05$ as statistically significant. Analyses were performed using R Statistical Software (version 4.1.3; R Foundation) and RStudio statistical software (version 2022.02.3 for Windows).

Ethical Considerations

This study was approved by the Institutional Review Board of the University of Wisconsin (Protocol Institutional Review Board ID 2017 - 0930). This study was conducted according to the guidelines established in the Declaration of Helsinki for studies involving human subjects. Written informed consent was obtained from all participants in the in-depth interviews, and online consent before the access to the survey in the Qualtrics platform. Students who completed the qualitative interview received US $20 on their campus card. Participants who completed both sets of surveys received US $10 on their student campus card. Data was de-identified for anaylsis.

Results

Cognitive Interviews

Table S1 in Multimedia Appendix 1 presents the characteristics of the students who participated in the interviews. The sample included 8 students, aged 19 to 22 years, representing diverse ethnicities, genders, and undergraduate years. The interview duration ranged from 20 minutes to 1 hour.

Overall, participants found most questions easy to answer and acceptable, with their interpretations “in their own words” aligning with the intended purpose of each question. However, for a few questions, 1 or more participants suggested that the interpretation could be improved by incorporating definitions, clarifications, or more specific terminology. Revisions based on these interviews are detailed in Table S2 in Multimedia Appendix 1.

The first version to assess the psychometric properties consists of 31 items (19 core questions+12 skip questions). Additionally, the survey includes 12 questions about sociodemographic characteristics.

Assessment of Psychometric Properties

A total of 181 students completed the first administration of the questionnaire between June and November 2022. Of these, 116 also completed the second administration. Six observations were excluded from the analysis due to missing answers in >50% of the questions (4 from the first administration and 2 from the second administration). All of them were from the group of students of color. Thus, 177 responses were available for the analysis of internal consistency, 115 for test-retest reliability assessment and 98 for construct validity.

Participants’ characteristics for each substudy are shown below in Table 1. Students of color represented 48.6% (86/177), 43.5% (50/115), and 57.1% (56/98) in substudies of assessment of internal consistency, test-retest assessment, and construct validity, respectively. Most of the students of color were Asian or Asian American, followed by those of Hispanic, Latinx, or Spanish origin. There were students representing all stages of academic studies. A larger number of students lived off campus, and the most frequent current financial situation was described as “tight but doing fine.”

XSL•FO
RenderX

Table . Demographic characteristics of the participants, the Color of Drinking validation study, University of Wisconsin-Madison, 2022.

Characteristics	Internal consistency (n 177)	Test-retest (n 115)	Construct (n 98)
Ethnicity[a]			
American Indian or Alaskan Native	3	3	3
Asian or Asian American	75	45	50
Black or African American	7	5	6
Hispanic, Latinx, or Spanish	14	7	11
Middle Eastern or North African	2	1	1
White	91	65	42
Biracial	3	2	3
Multiracial	1	2	1
Other identity	7	2	3
Gender			
Female	118	86	62
Male	50	23	30
Gender diverse	9	6	6
Self-identification			
Disabled person	34	21	22
Able-bodied or not disabled	117	79	66
Prefer not to answer	26	15	10
Stage of academic studies			
1st year undergraduate	20	8	17
2nd year undergraduate	42	25	30
3rd year undergraduate	48	30	20
4th year undergraduate	42	36	27
5th year undergraduate or other	16	15	4
Prefer not to answer	9	1	0
Living arrangement			
On campus	43	21	31
Off campus	122	92	66
Other	2	0	1
Prefer not to answer	10	2	0
Current financial situation			
It’s a financial struggle	22	13	10
It’s tight but I’m doing fine	80	53	54
Finances aren’t really a problem	65	47	34
Prefer not to answer	10	2	0
College or school			
College of Agricultural and Life Sciences (CALS)	28	21	14
College of Engineering	21	14	9
College of Letters and Science (L&S)	89	59	56
Wisconsin School of Business	16	11	11

XSL•FO
RenderX

Characteristics	Internal consistency (n 177)	Test-retest (n 115)	Construct (n 98)
School of Education	13	9	7
School of Human Ecology	4	3	1
School of Nursing	5	4	4
School of Pharmacy	4	3	2
Nelson Institute for Environmental Studies	3	2	1
School of Veterinary Medicine	1	1	0
School or college not listed	4	3	1
Prefer not to answer	9	2	0

[a]Multiple answer options were allowed.

Internal Consistency

The four dimensions evaluated presented good internal consistency, with Cronbach α coefficients ranging from 0.723 to 0.898 (Table 2).

XSL•FO
RenderX

Table . Internal consistency of 4 sets of items: impact of alcohol consumption on academics, the impact of microaggressions, witnessing microaggressions and alcohol intoxication, and bystander interventions on alcohol intoxication: Color of Drinking validation study, UW-Madison[a], 2022.

Dimension, Q[b], and items	Cronbach α coefficient (95% CI)	Cronbach α coefficient if an item is dropped
Impact of alcohol consumption on academics (n=177)	0.795 (0.689-0.854)	
Q8 How often have you experienced the following during the current semester?[c]		
8.1 I have been too hungover to attend class		0.74
8.2 I chose to drink instead of study		0.77
8.3 I received a poor grade because I chose to drink instead of study		0.78
8.4 Missed a class because of alcohol use		0.74
8.5 Performed poorly on an assignment due to alcohol use		0.74
Impact of microaggressions (n=22)	0.898 (0.832-0.94)	
Q17. How much did the microaggressions impact your …?[d]		
17.1 Sense of belonging at UW-Madison?		0.88
17.2 Health and wellbeing?		0.83
17.3 Ability to learn at UW-Madison?		0.85
Witnessing microaggressions and alcohol intoxication (n=12)	0.723 (0.643-0.782)	
Q25. Have you witnessed any of the following?[e]		
25.1 A friend makes a discriminatory statement to a stranger		0.72
25.2 A stranger makes a discriminatory comment to a friend		0.68
25.3 Your alcohol-intoxicated friend makes a discriminatory comment to a stranger		0.7
25.4 An alcohol-intoxicated stranger makes a discriminatory comment to a friend		0.68
25.5 A friend is passed out in a public space and unconscious due to alcohol		0.72
25.6 A stranger is passed out in a public space and unconscious due to alcohol		0.67
25.7 A friend being transported to detox because of alcohol intoxication		0.7
25.8 A stranger being transported to detox because of alcohol intoxication		0.7
Bystanders' interventions (n 112)	0.724 (0.649-0.779)	
Q26. Would you intervene in the following situations?[e]		
26.1 A friend is passed out in a public space and unconscious due to alcohol		0.67
26.2 A stranger is passed out in a public space and unconscious due to alcohol		0.68
26.3 A friend being transported to detox because of alcohol intoxication		0.65
26.4 A stranger being transported to detox because of alcohol intoxication		0.64

XSL•FO
RenderX

[a]UW-Madison: University of Wisconsin-Madison.

[b]Q: question.

[c]Answer categories: never, rarely, sometimes, often, and most of the time

[d]Answer categories: not at all, slightly, somewhat, quite a bit, and a great deal.

[e]Answer categories: yes and no.

Test-Retest Reliability

Approximately 33% (38/115) and 30% (35/115) of the students reported that they do not drink alcohol or did not drink during the last 30 days in test and retest surveys, respectively. The mean number of drinking days within the last 2 weeks was 1.4 and 1.8 for test and retest, respectively. In summary, the section "alcohol use" showed moderate to substantial reliability (Table S3 in Multimedia Appendix 1).

On test and retest surveys, 9.6% (11/114) of the participants reported that the alcohol culture at UW-Madison had impacted their academics (Table S4 in Multimedia Appendix 1). Most items on the section "alcohol culture and academics" showed moderate to substantial test-retest reliability, but 1 of them showed low test-retest reliability.

On test and retest surveys, 51.8% (59/114) of the participants reported that during the semester they had avoided specific areas on or off campus due to concerns about alcohol use by others. The proportion in which participants were negatively impacted by other students' alcohol consumption was 35% (40/113) for the test and 28% (32/113) for the retest. Items from these 2 sections, "areas avoided" and "impact of by other students' alcohol consumption" also showed mostly moderate to substantial test-retest reliability (Table S2 in Multimedia Appendix 1).

Among the students of color, 47% (29/62) and 40% (25/62) reported having experienced a microaggression on campus or surrounding areas on test and retest surveys. On test and retest surveys, 71% (81/114) and 73% (83/114) of the students had witnessed at least once a microaggression on campus or surrounding areas during their time at UW-Madison. A summary of test-retest reliability assessment for experiencing, witnessing, and bystander intervention of microaggressions and alcohol intoxications is presented in Table S4 in Multimedia Appendix 1. Most of the items from these dimensions showed moderate to substantial test-retest reliability.

From test and retest surveys, 33% (37/112) and 25% (28/112) of the participants answered that their own or alcohol consumption by other students impacted their health during the time they attended UW-Madison. At test and retest, 43% (49/113) and 42% (47/113) of the students reported that the alcohol culture had impacted their overall sense of belonging at UW-Madison. Most of the items from the section "impact on health and sense of belonging" showed moderate to substantial test-retest reliability, except for 1 related to the impact on health (Table S4 in Multimedia Appendix 1).

Construct Validity

Students with AUDIT scores ≥7 reported a higher number of drinking days within the last 2 weeks than those with lower AUDIT scores. A similar association was found regarding the maximum number of drinks in a day and the AUDIT score with a cutoff point of 7 (Table 3). Correlations between the number of drinking days, the maximum number of drinks in a day and the AUDIT score as a continuous variable were moderate to high, r=0.630 (95% CI 0.533-0.719) and r=0.647 (95% CI 0.548-0.741), respectively (Table 3).

A few students reported that they received a poor grade because they had chosen to drink instead of study. Their GPA was lower than the other students' (Table 4).

Table . Construct validity. Drinking days within the last 2 weeks and a maximum number of drinks in a day among participants with AUDIT[a] scores less than and greater than or equal to 7. CoD[b] validation study, University of Wisconsin-Madison, 2022.

	AUDIT score		
	<7 (n=62)	≥7 (n=35)	
Variables from the Color of Drinking questionnaire	Median(IQR)	Median (IQR)	*P* value[c]
Drinking days within the last 2 wk	0 (0 - 2)	3 (0 - 5)	<.0001
Maximum number of drinks in a day	0 (0 - 3)	6 (3 - 10)	<.0001

[a]AUDIT: Alcohol Use Disorders Identification Test.

[b]CoD: Color of Drinking.

[c]Wilcoxon rank sum test.

XSL•FO
RenderX

Table . Construct validity. Self-reported grade point average (GPA[a]), microaggression subscale from the REMS[b] and SBI[c] according to selected items from the CoD[d] questionnaire validation study. UW-Madison[e], 2022.

Dimension: CoD questionnaire item	Variable (other instrument)	N	Mean (SD)	Median (IQR)	*P* value[f]
Dimension: alcohol consumption and academics					
CoD questionnaire, item 8.3 ("I received a poor grade because I chose to drink instead of study")	GPA, accumulated				.04
At least once during the current semester		7	3.19 (0.57)	3.3 (2.7 - 3.68)	
Never		71	3.6 (0.37)	3.7 (3.4 - 3.9)	
	GPA, last term				.02
At least once during the current semester		6	2.9 (0.77)	2.9 (2.28 - 3.55)	
Never		70	3.59 (0.58)	3.8 (3.5 - 4)	
Dimension: microaggressions					
CoD questionnaire, Q13. "In your time at UW-Madison, have you experienced a microaggression on campus or surrounding areas?"	Microaggressions subscale from REMS[g]				<.001
Yes		27	2.04 (1.83)	2 (0 - 3.5)	
No		40	0.68 (1.4)	0 (0 - 1)	
Dimension: impact on the sense of belonging					
CoD questionnaire, Q28. "Has the alcohol culture impacted your overall sense of belonging at UW-Madison?"	SBI[h]				<.001
Yes		41	16.07 (2.23)	16 (15-18)	
No		57	18.04 (2.15)	18 (17 - 19)	

[a]GPA: grade point average.

[b]REMS: Racial and Ethnic Microaggression Scale.

[c]SBI: Sense of Belonging Index.

[d]CoD: Color of Drinking.

[e]UW-Madison: University of Wisconsin-Madison.

[f]Wilcoxon rank sum test (*P* value).

[g]Score ranges from 0 to 5.

[h]Score ranges from 5 to 30.

Additionally, 27 of 67 students of color answered that in their time at UW-Madison they had experienced a microaggression on campus or surrounding areas. They scored higher on the REMS than the rest of the students.

Lastly, 41 of 98 students answered that the alcohol culture impacted their overall sense of belonging at UW-Madison. These students scored lower in the SBI in comparison with the rest of the students (Table 4).

Redesign and Assessment for Questions With Low Performance and Developing the Final Questionnaire

The "health impacts" section showed a low test-retest reliability. When the qualitative responses were evaluated, students brought

XSL•FO
RenderX

up different interpretations but also very profound answers and concepts, such as mental health struggles or experiencing sexual assault. Based on these results, the decision was made to rewrite and retest the validity of this question grounded on the evaluation of content validity, and other series of in-depth interviews were used to generate a revised set of questions. Eight items were developed for the personal consumption of alcohol on diverse aspects of health, and the same number of items were included for the consumption of others.

Online surveys were conducted including the 16 new items with another sample of undergraduate students selected following the same procedures. The sample included 118 students who completed the first questionnaire. Most participants were women (84, 71%), and 66 participants were self-identified as White. Sixty-seven participants also completed the second questionnaire.

The results for internal consistency and test-retest reliability were Cronbach α values of .735 and .855, for each set of questions, respectively, and most of the items presented κ coefficients from 0.476 to 0.877 (Tables S5 and S6 in Multimedia Appendix 1).

Discussion

The validation process outlined here included the assessment of reliability and validity of the Color of Drinking questionnaire, used in an exploratory study of the impact of UW-M's alcohol culture on students of color. The process of validating the original questionnaire resulted in a revised instrument, that showed good internal consistency, acceptable to good test retest reliability for most items and good construct validity.

In the alcohol and academics dimension, we found that questions related to personal academic consequences had higher internal consistency and a higher Cronbach α than the other questions in the same dimension. Questions about the secondhand negative academic consequences were found to be highly relevant but not highly correlated with each other. The question "my professors or TAs invited me out to places where alcohol is served" tested poorly when placed with the other secondhand academic consequences. For the finalized questionnaire we decided to separate the personal academic consequences, and secondhand negative consequences as 2 separate subdomains. We also removed the aforementioned question due to some unclearness about the scope and meaning of the item identified during the qualitative interview and little consistency with other items in the expected subdomain.

In the bystander intervention dimension, there were issues with test-retest reliability among 4 questions. Bystander discrimination questions from the questionnaire had low test-retest reliability. The questions were more prone to variation over time than the other bystander questions. The decision was made to remove these 4 questions from the finalized version of the questionnaire. Other validation processes in college settings show adequate validity and reliability when assessing other types of bystander interventions [24].

The health impacts section showed a low test-retest reliability, however based on responses there was important information to include in this section. Thus, the section was rewritten and tested, reaching good reliability.

A questionnaire designed for populational health studies should capture what it was meant to collect when designed in a reliable manner [25,26]. This validation process will not only allow us to measure what we want to measure but also will make our data comparable over time and with other college settings [27]
.

This study has limitations. Although there was a random selection of students invited to participate, we recognize that those more sensitive to racial issues may have been more likely to take part in this study. This potential selection bias could influence the generalizability of our findings to the broader university population.

In conclusion, the process of validating the original questionnaire for the Color of Drinking study resulted in a revised instrument, that showed good internal consistency, acceptable to good test-retest reliability for most items and good construct validity.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional materials: additional tables S1-S6.
[DOCX File, 45 KB - i-jmr_v14i1e64720_app1.docx]

References

1. Harmful and underage college drinking. National Institute on Alcohol Abuse and Alcoholism. 2023. URL: https://www.niaaa.nih.gov/publications/brochures-and-fact-sheets/college-drinking [accessed 2025-03-20]
2. Lorant V, Nicaise P, Soto VE, d'Hoore W. Alcohol drinking among college students: college responsibility for personal troubles. BMC Public Health 2013 Jun 28;13:615. [doi: 10.1186/1471-2458-13-615] [Medline: 23805939]
3. American College Health Association. American College Health Association-National College Health Assessment II: Undergraduate Student Reference Group DataReport Spring 2019.: American College Health Association; 2019 URL:

XSL•FO
RenderX

https://www.acha.org/wp-content/uploads/2024/07/NCHA-II_SPRING_2019_UNDERGRADUATE_REFERENCE_GROUP_DATA_REPORT.pdf [accessed 2025-03-20]
4. Wechsler H, Nelson TF. What we have learned from the Harvard School Of Public Health College Alcohol Study: focusing attention on college student alcohol consumption and the environmental conditions that promote it. J Stud Alcohol Drugs 2008 Jul;69(4):481-490. [doi: 10.15288/jsad.2008.69.481] [Medline: 18612562]
5. Satinsky S, Washington RL, Pastor J, Wagner AK. Campus high-risk drinking culture as a social justice issue: a commentary on the potential impact on the mental health and well-being of marginalized college students. Am J Health Educ 2017 Jul 4;48(4):222-225. [doi: 10.1080/19325037.2017.1316695]
6. Washington R, Marconi A, Reeves M, Jardas E. The color of drinking: an exploratory study of the impact of UW–madison's alcohol culture on students of color. University Health Services. 2020. URL: https://www.uhs.wisc.edu/wp-content/uploads/2020/01/10_1_2018_ColorOfDrinkingBooklet_RLG.pdf [accessed 2025-03-20]
7. Marconi A, Washington R, Reeves M, Bradley Q, Ayala A, Griggs C. Examining racial microaggressions and alcohol use among marginalized populations at a predominately White institution. J Am Coll Health 2024;72(6):1886-1895. [doi: 10.1080/07448481.2022.2098027] [Medline: 35930365]
8. Rosciglione A. If you don't drink, you don't belong': a deep dive into UW's color of drinking survey. The Daily Cardinal. 2022. URL: https://www.dailycardinal.com/article/2022/11/if-you-dont-drink-you-dont-belong-a-deep-dive-into-uws-color-of-drinking-survey [accessed 2025-03-20]
9. Nadal KL. The Racial and Ethnic Microaggressions Scale (REMS): construction, reliability, and validity. J Couns Psychol 2011 Oct;58(4):470-480. [doi: 10.1037/a0025193] [Medline: 21875180]
10. Forrest-Bank S, Jenson J. Differences in experiences of racial and ethnic microaggression among Asian, Latino/Hispanic, Black, and White young adults. J Sociology Soc Welfare 2015 Jan 1;42(1). [doi: 10.15453/0191-5096.3891]
11. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016 Jun;15(2):155-163. [doi: 10.1016/j.jcm.2016.02.012] [Medline: 27330520]
12. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Statist Med 1998 Jan 15;17(1):101-110. [doi: 10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.3.CO;2-5]
13. Streiner DL, Norman GR. Health Measurement Scales: Oxford University Press; 2008. [doi: 10.1093/acprof:oso/9780199231881.001.0001]
14. Sue DW, Capodilupo CM, Holder AMB. Racial microaggressions in the life experience of Black Americans. Prof Psychol: Res Pract 2008;39(3):329-336. [doi: 10.1037/0735-7028.39.3.329]
15. Diener E, Wirtz D, Tov W, et al. New well-being measures: short scales to assess flourishing and positive and negative feelings. Soc Indic Res 2010 Jun;97(2):143-156. [doi: 10.1007/s11205-009-9493-y]
16. Beatty PC, Willis GB. Research Synthesis: The Practice of Cognitive Interviewing. Public Opin Q 2007;71(2):287-311. [doi: 10.1093/poq/nfm006]
17. National Research Council. Cognitive Aspects of Survey Methodology: National Academies Press; 1984. [doi: 10.17226/930ISBN:978-0-309-07784-2]
18. Willis GB. Analysis of the Cognitive Interview in Questionnaire Design: Oxford University Press; 2015. URL: https://global.oup.com/academic/product/analysis-of-the-cognitive-interview-in-questionnaire-design-9780199957750?cc=ar&lang=en&amp [accessed 2025-03-29]
19. 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]
20. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. Addiction 1993 Jun;88(6):791-804. [doi: 10.1111/j.1360-0443.1993.tb02093.x] [Medline: 8329970]
21. Pedler ML, Willis R, Nieuwoudt JE. A sense of belonging at university: student retention, motivation and enjoyment. J Further Higher Educ 2022 Mar 16;46(3):397-408. [doi: 10.1080/0309877X.2021.1955844]
22. Bador TF, Higgins-Biddle JC, Saunders JB, Monteiro MG, Salud OM. AUDIT: cuestionario de identificación de los transtornos debidos al consumo de alcohol: pautas para su utilización en atención primaria. : World Health Organization; 2001 URL: https://apps.who.int/iris/handle/10665/331321 [accessed 2025-03-20]
23. Nunnally JC. Psychometric Theory, 2nd edition: McGraw-Hill; 1978.
24. Banyard VL, Plante EG, Moynihan MM. Rape prevention through bystander education: bringing a broader community perspective to sexual violence prevention. : NCJRS; 2005 URL: https://www.ncjrs.gov/pdffiles1/nij/grants/208701.pdf [accessed 2025-03-20]
25. Taherdoost H. Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research: SSRN Electron J Elsevier BV; 2016. [doi: 10.2139/ssrn.3205040]
26. Aithal A, Aithal PS. Development and validation of survey questionnaire & experimental data – a systematical review-based statistical approach. IJMTS 2020;5(2):233-251. [doi: 10.47992/IJMTS.2581.6012.0116]
27. Kazi AM, Khalid W. Questionnaire designing and validation. J Pak Med Assoc 2012 May;62(5):514-516 [FREE Full text] [Medline: 22755326]

XSL•FO
RenderX

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test
REMS: Racial and Ethnic Microaggression Scale
SBI: Sense of Belonging Index
UW-Madison: University of Wisconsin-Madison

Edited by TDA Cardoso; submitted 24.07.24; peer-reviewed by A Ortega, Y Pardo; revised version received 11.02.25; accepted 19.02.25; published 07.04.25.

Please cite as:
Marconi A, Washington R, Jovaag A, Blomme C, Knobeloch A, Irazola V, Muros Cortés C, Gutierrez L, Elorriaga N
The Color of Drinking Survey Questionnaire for Measuring the Secondhand Impacts of High-Risk Drinking in College Settings: Validation Study
Interact J Med Res 2025;14:e64720
URL: https://www.i-jmr.org/2025/1/e64720
doi:10.2196/64720

© Agustina Marconi, Reonda Washington, Amanda Jovaag, Courtney Blomme, Ashley Knobeloch, Vilma Irazola, Carolina Muros Cortés, Laura Gutierrez, Natalia Elorriaga. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 7.4.2025. 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

Evaluating the Effect of the JUUL2 System With 5 Flavors on Cigarette Smoking and Tobacco Product Use Behaviors Among Adults Who Smoke Cigarettes: 6-Week Actual Use Study

Nicholas I Goldenson[1], PhD; Saul Shiffman[2], PhD; Mark A Sembower[2], MS; Arielle Selya[2], PhD; Steve Pype[2], BS; Ryan A Black[1], PhD

[1]Juul Labs, Inc, Washington, DC, United States

[2]PinneyAssociates, Inc, Pittsburgh, PA, United States

Corresponding Author:
Nicholas I Goldenson, PhD
Juul Labs, Inc
1000 F Street NW
Suite 800
Washington, DC, 20004
United States
Phone: 1 510 501 5964
Email: nicholas.goldenson@juul.com

Abstract

Background: Adults who switch completely from smoking cigarettes to using electronic nicotine delivery systems (ENDS) substantially reduce their exposure to toxicants and carcinogens that are associated with smoking-related diseases.

Objective: This 6-week actual use study—a prospective uncontrolled real-world study designed to evaluate quasi-naturalistic product use—aimed to assess switching behavior among US adults who smoked cigarettes and were provided with JUUL2 ENDS products.

Methods: US adults who smoked cigarettes every day but were predominantly not ready to quit (N=1160; mean age 39.42, SD 11.03 years; 641/1160, 55.26% female participants; 667/1160, 57.5% non-Hispanic White; mean cigarettes per day 14.11, SD 8.96; only 1% [11/1160] planning to stop smoking within 30 days; and 481/1160, 41.47% dual users) were recruited to use JUUL2 ENDS products (18 mg/mL nicotine) in 1 of 5 flavors in real-world environments for 6 weeks. Participants who expressed sufficient interest in using JUUL2 products were enrolled at 24 different consumer research sites across the United States into one of the two following study arms: (1) traditional flavors (Virginia Tobacco and Polar Menthol, 10 sites); or (2) complex flavors (Autumn Tobacco, Summer Menthol, and Ruby Menthol, 14 sites). No instructions regarding JUUL2 product use or cigarette smoking were provided. After a 1-week trial period, participants were provided with their preferred flavor for 6 weeks of ad libitum use (10 pods per week). In total, 6 weekly web-based surveys were used to assess switching (smoking abstinence) and smoking reduction; dependence and respiratory symptoms were assessed at baseline and week 6.

Results: Across the 5 flavor groups at week 6, the rates of complete past-7-day switching away from cigarettes ranged from 38.2% (79/207) to 47.3% (95/201), and 24.3% (55/226) to 33.9% (74/218) of participants reported complete past-30-day switching. Participants who used the 3 menthol-flavored (vs 2 tobacco-flavored) JUUL2 products had significantly higher rates of past-30-day switching at week 6 (odds ratio 1.36, 95% CI 1.04-1.78). Compared to their baseline values when they were smoking, the past-30-day switchers at week 6 had significantly reduced their dependence (mean differences in dependence, cigarettes – JUUL2: 0.57-0.99; $P<.001$) and self-reported frequency of respiratory symptoms ($P<.05$). Among participants who continued to smoke at week 6, 50.9% (59/116) to 62.9% (73/116) reduced their daily cigarette consumption by at least 50% from baseline.

Conclusions: Adoption of JUUL2 ENDS products can likely help substantial proportions of US adults who smoke to switch completely away from cigarettes or meaningfully reduce their cigarette consumption, thereby reducing their dependence on tobacco products and improving their respiratory symptoms.

(Interact J Med Res 2025;14:e60620) doi:10.2196/60620

XSL•FO
RenderX

KEYWORDS

cigarette; smoking; electronic nicotine delivery system; switching; actual use study; real world

Introduction

Background

Evidence demonstrates that adults who switch completely from smoking cigarettes to using electronic nicotine delivery systems (ENDS) substantially reduce their exposure to toxicants and carcinogens that are associated with smoking-related diseases [1-4]. Accordingly, the primary public health benefit of ENDS is derived from facilitating switching behavior among adults who smoke.

For ENDS products that are currently marketed in the United States, observational studies and randomized trials have assessed switching behavior among adults who smoke [5-9]. However, for premarket ENDS products (ie, those not currently marketed in the United States), evaluating naturalistic product use behaviors and transitions over time in real-world settings requires providing the premarket product to adults who smoke cigarettes. Actual use studies—prospective uncontrolled real-world studies designed to evaluate quasi-naturalistic product use—are the primary method for assessing the use of premarket tobacco products under ecologically valid conditions. Actual use studies were initially developed for pharmaceuticals to determine whether prescription drugs could be switched to over-the-counter sale by assessing use and safety of the drug in real-world settings designed to mimic over-the-counter use [10,11].

In recent years actual use studies have evaluated use of tobacco products, including oral nicotine pouches [12,13] and heated tobacco (heat-not-burn) products [14,15], among adults who smoke to prospectively assess changes in cigarette smoking following quasi-naturalistic use of the new tobacco products in real-world settings. In contrast to randomized cessation trials, actual use studies do not require participants to have active plans to quit smoking to qualify, do not set explicit goals for product use or smoking, and do not provide instructions, encouragement, or behavioral support to quit. Thus, actual use studies are particularly well-suited to evaluate the ability of ENDS to facilitate switching behavior (smoking abstinence) following periods of unguided self-determined use.

The JUUL2 system is a next-generation pod-based ENDS product that is designed to facilitate high levels of complete switching away from cigarettes among adults who smoke cigarettes. Pharmacokinetic studies show that JUUL2 18 mg/mL (1.5% nicotine by weight) products can deliver similar levels of nicotine as ENDS products with higher nicotine concentrations due to increased aerosol production [16], and behavioral pharmacology data demonstrate that Polar Menthol–flavored 18 mg/mL JUUL2 pods are rated as significantly more satisfying than menthol-flavored pod-based ENDS with 5% nicotine concentrations [17]. The JUUL2 system is currently marketed in the United Kingdom; however, there is a lack of real-world data on the switching potential of these JUUL2 products among US adults who smoke.

Objectives

The primary aim of this actual use study was to evaluate complete switching away from cigarettes (ie, past-7-day and past-30-day smoking abstinence) among US adults who smoked cigarettes every day and were provided with JUUL2 products in their preferred flavor for 6 weeks. Secondary aims of the study were to assess: (1) changes in dependence and respiratory symptoms among participants who switched completely away from smoking at week 6; (2) changes in cigarette consumption among participants who continued smoking at week 6; (3) patterns of JUUL2 product use and subjective responses to JUUL2 products; and (4) factors associated with switching behavior, including JUUL2 study product flavor (menthol vs tobacco), cigarette flavor of usual brand (menthol vs nonmenthol), and baseline ENDS use (dual use along with cigarette smoking).

Methods

Design

The study was designed as an actual use study—a prospective real-world study intended to simulate quasi-naturalistic product use—in which US adults who smoked cigarettes were provided with JUUL2 study products in participant-selected flavors for 6 weeks of ad libitum use in their natural environment. All potential participants first completed an initial prescreening survey (in person or via telephone) and then the formal recruitment screening on the web. The study had 2 distinct arms, implemented at separate research sites with different geographic catchment areas that tested 1 of the 2 following flavor series: (1) traditional flavors (Virginia Tobacco and Polar Menthol), which correspond to ENDS flavors that have been authorized for marketing by US Food and Drug Administration (FDA; 10 sites); or (2) complex flavors (Autumn Tobacco, Summer Menthol, and Ruby Menthol), which represent adult-oriented flavor mixtures that have not yet received market authorization from FDA (14 sites). The 2 arms were conducted at distinct sites and did not share participant pools to evaluate traditional and complex flavors independently. The decision to separate the traditional and complex flavors was an a priori design based on differences in the composition of traditional and complex flavors (see Study Products section), with implied different regulatory positions (ie, FDA has authorized tobacco- and menthol-flavored ENDS, but has not yet authorized nontobacco or nonmenthol flavors or blends), and to avoid the complexity of each participant and site testing all 5 flavors. Study design and procedures for the 2 arms were identical, and the geographic and demographic position of the research sites were roughly parallel, but sites recruited participants from their local catchment area, thus there was no overlap in potential participants across study sites. Importantly, participants were not assigned to sites or arm by the study (randomly or otherwise) but were recruited from each site's distinct geographic catchment area. The study sponsor did not play any role in the recruitment of participants and individuals were primarily

XSL•FO
RenderX

recruited from existing databases maintained by the study sites or in person (not sponsor databases).

Eligible participants first completed a 1-week trial period to determine which flavor they would exclusively use for the 6-week actual use period. During the 6-week actual use period, participants completed weekly web-based survey assessments (Multimedia Appendix 1).

At prescreening, participants were informed that the research study involved using new electronic cigarette products for 7 weeks and that participation in the study did not require stopping smoking cigarettes—the study was not framed or described as a study of behavior change or smoking cessation. During the trial week and 6-week actual use period, no goals or instructions were provided regarding cigarette smoking; participants could smoke ad libitum throughout the study, or reduce or forego smoking, at their discretion. Similarly, no instructions (beyond the user guide included in the product package) were provided regarding use of the JUUL2 study products with respect to amount or persistence of use. The only directions were those that come with marketed products, describing how to use the product and how to charge the device.

Participants

The study sample included healthy non–treatment-seeking English-speaking US adults (aged 22 to 65 years) who smoked cigarettes every day and lived close to one of the 24 shopping-mall–based study sites distributed across 15 states (Multimedia Appendix 2). Eligibility criteria were as follows: (1) smoked ≥100 combustible tobacco cigarettes in their lifetime and smoked cigarettes for ≥12 months before screening; (2) smoked cigarettes every day and smoked an average of ≥5 cigarettes per day; (3) interested in trying and using JUUL ENDS products; (4) access to a computer or laptop, smartphone, or tablet with internet access; (5) provided consent for participation and acknowledged willingness and ability to comply with all study requirements. Interest in stopping smoking was not an eligibility criterion.

Exclusion criteria included the following: (1) use of any nicotine replacement therapy (eg, patch and gum) or prescription smoking cessation medications (eg, varenicline and bupropion) within 30 days before screening; (2) diagnosis of a clinically significant medical (eg, asthma, chronic obstructive pulmonary disease, heart disease, high blood pressure, and cancer) or psychiatric disease; (3) was pregnant, nursing, or intended to become pregnant at any time through the end of study; (4) participated in a research study about tobacco products or ENDS within the past 30 days; (5) was a current or former employee or related to a current or former employee of the tobacco or ENDS industry or any vendor associated with conduct of the study; and (6) operated or lived with family member who operated any home childcare or health care services. There were no eligibility criteria regarding prior use of ENDS; participants could be ENDS naive, former ENDS users, or current dual users. Recruitment aimed to enroll a diverse sample with respect to sociodemographic and tobacco product use characteristics, according to the following soft quotas: (1) ≤60% of male or female participants; (2) at least 35% of participants who identified as non-Hispanic White; (3) ≥50% participants aged <45 years; (4) ≥40% of participants who smoked mentholated cigarettes; and (5) ≥50% of participants who had ever used ENDS. Individuals were primarily recruited from existing databases maintained by the study sites and the study sponsor did not play any role in recruitment of participants.

Ethical Considerations

Participants provided written informed consent. All surveys were completed remotely on the web. Participants were compensated US $20 for completing the baseline assessment, US $15 for completing the product trial assessment, and US $25 for completing each weekly assessment. Additional compensation was provided for completing all study assessments and returning study products, with maximum total compensation of US $450. The Advarra Institutional Review Board approved the study protocol (Pro00068019).

Procedure

Following enrollment, participants completed the baseline survey on the web and visited their study site, where they were provided with a JUUL2 device, charging dock, and 2 pods of each of the flavors in their respective study arm (2 traditional flavors or 3 complex flavors) for a 1-week product trial period. Participants were instructed to use each flavor for at least one day, enough to form an opinion of the product; no other instructions were provided. After the trial period, participants completed a brief web-based survey in which they rated their interest in each of the trialed JUUL2 pod flavors. These responses were used to determine if participants qualified for the 6-week actual use period and which flavor they would use (see Interest in Using and Purchasing JUUL2 Products subsection of Measures).

At the start of week 1, participants received an additional JUUL2 device. Every 2 weeks (at the start of week 1, week 3, and week 5) participants were provided with 20 pods for use during the upcoming 2 weeks (10 pods per week, 60 pods total). Participants were invited by email to complete assessments at the end of each of the study weeks and received compensation upon completion of each weekly assessment. Participants had 3 days from the initial invitation to complete the weekly surveys and they received multiple reminder emails prompting them to complete the assessment (surveys not completed within this window were considered missing).

Participants were instructed in the informed consent form and by study staff to inform study staff of any adverse events (AEs) and intercurrent illnesses experienced during the study. In addition, a specific inquiry regarding AEs (ie, if participants had experienced any changes to their health [yes or no]) was made at the product trial (week 0), week 3, and week 6 surveys; at the week 1, 2, 4, and 5, surveys participants were prompted to report any changes to their health to the study staff. Reported health events were assessed and classified via clinical interviews by trained medical personnel.

Study Products

The JUUL2 system is closed-system pod-based ENDS that is inhalation-actuated and does not have any user-modifiable settings, controls, or buttons. Preclinical studies demonstrate

XSL•FO
RenderX

that the aerosol produced by the JUUL2 system contains substantially lower levels of harmful and potentially harmful constituents than cigarette smoke [18]. The 5 JUUL2 pod flavors evaluated in this study are all currently marketed in the United Kingdom. Virginia Tobacco and Polar Menthol have traditional tobacco and menthol flavors, respectively. The 3 complex flavors (Autumn Tobacco, Summer Menthol, and Ruby Menthol) consist of a tobacco or menthol base, respectively, with top notes of fruit: apple for Autumn Tobacco, tropical fruit for Summer Menthol, and berry for Ruby Menthol.

Measures

Participants Sociodemographic and Tobacco Use Characteristics

Sociodemographic and tobacco or nicotine product use characteristics were assessed in the recruitment screener and baseline assessment with items and measures adapted from national surveys, with some modified for relevance.

Interest in Using and Purchasing JUUL2 Products

At the end of the trial week, participants were asked 2 questions about each of the JUUL2 products they used as follows: (1) "How interested are you in using the JUUL2 study product for the next 6 weeks?" (5-point response scale, from 1 ["not at all interested"] to 5 ["extremely interested"]) and (2) "If the JUUL2 study products you tried this week were available for purchase, would you buy one or more of the products?" (4-point response scale, from 1 ["definitely would not buy"] to 4 ["definitely would buy"]). Only participants who responded 3 or higher to both items for at least one flavor continued to the 6-week actual use period. This was intended to mirror real-world self-selection for product use, as individuals who were not interested in using or purchasing a specific flavor of a consumer product would be unlikely to purchase and adopt it. Participants were provided with the single flavor they expressed the greatest interest in using for the 6-week period; those who expressed exactly equivalent interest in multiple flavors were assigned to the flavor group in their respective study arm that was most in need of enrollees, which helped balance the sample size of flavor groups.

Cigarette Smoking

At each weekly survey, participants were asked if they had smoked cigarettes (even one puff) in the past 7 days (yes or no). At the week 6, survey participants were asked if they smoked in the past 30 days: "no" responses were operationalized as switching. Participants who reported smoking in the past 7 days at any survey between weeks 3 and 6 (ie, within 30 days of week 6) were coded as past-30-day smokers, even if they did not complete the week 6 survey or reported past-30-day abstinence at week 6.

Participants who reported smoking in the past 7 days also reported frequency and intensity of cigarette smoking during that period—daily cigarette consumption was calculated as cigarettes smoked per day in the past 7 days, including nonsmoking days (ie, product of the past-7-day frequency and daily intensity).

Dependence on Cigarettes and JUUL2 Products

Dependence on cigarettes was assessed at baseline and dependence on JUUL2 products was assessed at week 6 among participants who had not smoked in the past 30 days with the Tobacco Dependence Index (range 1-5, with higher scores indicating greater levels of dependence), a measure psychometrically validated in the Population Assessment of Tobacco and Health study for cross-product comparisons [19,20].

Respiratory Symptoms

Frequency of respiratory symptoms was assessed at baseline, when all participants smoked cigarettes every day, and at week 6 among the past-30-day switchers using the Respiratory Symptoms Experience Scale (range 1-5, with higher scores indicating greater frequency of respiratory symptoms), which was developed and validated for adults who smoke cigarettes but may not have clinical respiratory disease [21].

Patterns of JUUL2 Product Use

At each weekly survey, participants who reported using JUUL2 products in the past 7 days reported the number of days they used JUUL2 products and completed 2 items that assessed intensity of use as follows: (1) number of use episodes per day and (2) number of puffs per day.

Subjective Responses to JUUL2 Products

Subjective responses to JUUL2 products were assessed at week 6 with the modified Product Evaluation Scale [22].

Statistical Analysis

The study was prospectively powered for its primary aim, to evaluate rates of the past-30-day switching using the Clopper-Pearson exact method [23] with an estimated expected switch rate of 20% (based on switch rates observed in a prior observational study of US adults who purchased JUUL products), achieving a CI of 5%. Power analyses were not conducted for the secondary analyses.

The rate of past-7-day switching was calculated at each weekly survey within each of the 5 flavor groups; the rate of past-30-day switching was also evaluated at the week 6 Survey. A logistic regression model tested the association of JUUL2 flavor group and past-30-day switching at week 6, with Virginia Tobacco as the reference group; separate models contrasted the 2 combined tobacco flavors (Virginia and Autumn Tobacco) to the 3 combined menthol flavors (polar, summer and Ruby Menthol), and the 3 individual menthol flavors, respectively, to Virginia Tobacco. A repeated-measure logistic regression model included the past-7-day switching averaged across all 6 weeks as the dependent variable and JUUL2 flavor group, menthol cigarette preference (mentholated vs nonmentholated), and time (week) since baseline (coded as continuous, range 1-6) as simultaneous regressors—all 2-way interaction terms and the 3-way JUUL2 flavor × cigarette flavor × time interaction term were also included.

A separate repeated-measure logistic regression model was used to assess the association of baseline ENDS use (past 30 days vs not past 30 days) and the past-7-day switching. In addition,

XSL•FO
RenderX

mixed effects models evaluated linear changes over time in metrics of daily JUUL2 product use across the 6-week period.

Paired *t* tests were used to assess within-person changes in levels of dependence from baseline to week 6 (from cigarettes to JUUL2 products) among participants who reported no past-30-day smoking at week 6; dual users were not included in this analysis as they continued to smoke cigarettes and their dependence on JUUL2 could not be isolated. Changes in respiratory symptoms from baseline to week 6 were similarly assessed among complete 30-day switchers to assess whether switching affected their respiratory symptoms. Among participants who reported past-7-day smoking at week 6, changes in daily cigarette consumption (cigarettes per day on smoking days) from baseline were also tested with paired *t* tests. The proportion of participants reporting reductions of ≥50% was also calculated.

The primary analyses of switching used all observed (nonmissing) data without imputing smoking status to missed observations. This approach is typical for observational studies [24], in contrast to smoking cessation treatment trials, where there is concern that participants might avoid reporting smoking to avoid admitting to "failure" in an agreed behavioral goal [25]. To assess the potential for missingness to introduce bias, analyses were performed to compare participant baseline characteristics between participants who completed all 6 weekly surveys and those who missed any follow-ups. As a sensitivity analysis, the primary switching analyses were repeated while imputing smoking to all missing observations (intent-to-treat).

In addition, among participants who smoked mentholated cigarettes, differences in sociodemographic and tobacco use characteristics were compared between those who selected tobacco-flavored and menthol-flavored JUUL2 products.

Data were analyzed using SAS version 9.4 (SAS Institute Inc) with alpha level set to 0.05; no adjustments were made for multiple testing.

Results

Participant Accrual and Disposition

The enrollment flow in the traditional and complex flavor arms is provided in Multimedia Appendices 3 and 4, respectively. In the traditional flavors arm, 1874 adults were invited to complete the prescreener; 72.68% (1362/1874) completed it and were sent the eligibility screener; and 54.99% (749/1362) subsequently completed the recruitment screener and met all eligibility criteria. Of the eligible participants, 667 (89.1%) provided informed consent and were sent the baseline assessment (Multimedia Appendix 3). Approximately 99% (661/667) of those who started the baseline assessment completed it and were invited to participate in the trial week, 78.8% (521/661) subsequently completed the trial week assessment and were eligible to continue into the 6-week actual use period; 87% (575/661) completed the trial week survey; and 8.2% (54/661) were ineligible to continue into the 6-week actual use period based on expressing insufficient interest in using the study products.

In the complex flavors arm 2625 adults were invited to complete the prescreener; 71.7% (1882/2625) completed it and were sent the eligibility screener and 56.7% (1067/1882) subsequently completed the recruitment screener and met all eligibility criteria. Of the eligible participants, 942 (88.3%) provided informed consent and were sent the baseline assessment (Multimedia Appendix 4). Approximately 98% (927/942) of those who started the baseline assessment completed it and were invited to participate in the trial week, 79.9% (741/927) subsequently completed the trial week assessment and were eligible to continue into the 6-week actual use period; 13.6% (126/927) did not complete the trial week survey; and 6.5% (60/927) were ineligible to continue into the 6-week actual use period because they did not express sufficient interest in using the study products.

At screening, 61 (1.9%) individuals were excluded due to lack of interest in trying JUUL products. At the end of the trial week, the mean ratings of interest in purchasing JUUL2 products ranged from 3.37 to 3.44 on the 4-point response scale; ratings of interest in individual JUUL2 flavors ranged from 3.65 (Virginia Tobacco) to 4.12 (Polar Menthol) on the 5-point response scale (Multimedia Appendix 5). In the traditional flavors arm, 42.2% (242/574) of the participants expressed equivalent interest in both flavors (Virginia Tobacco and Polar Menthol), and in the complex flavors arm, 29.3% (232/792) expressed equivalent interest in all 3 flavors—27.3% (216/792) rated 2 of the 3 flavors equally and 43.4% (344/792) expressed a unique preference for a single flavor. There was no significant difference in the proportion of participants in the traditional (54/661, 8.2%) and complex (60/927, 6.5%) flavors arms that did not continue in the study due to lack of interest in using JUUL2 products at the end of the trial week (χ^2_1=1.67, *P*=.20). In the traditional flavors arm, 20.3% (106/521) of the participants met the threshold for inclusion for only one flavor. In the complex flavors arm, 30.6% (227/741) qualified for 1 or 2 flavors (ie, did not meet threshold for at least one flavor) and 8.8% (65/741) qualified for only one flavor.

The analytic sample consisted of 1160 total participants, roughly evenly divided into the 5 flavor groups: 242 participants in the Virginia Tobacco group, 239 in the Polar Menthol group, 219 in the Autumn Tobacco group, 236 in the Summer Menthol group and 224 in the Ruby Menthol group (Multimedia Appendices 3 and 4).

Sample Characteristics

Characteristics of the sample are reported for all participants (Table 1) and separately within each of the 5 individual flavor groups (Multimedia Appendix 6). Baseline characteristics were generally similar between JUUL2 flavor groups, although larger proportions of participants who selected menthol-flavored JUUL2 products smoked mentholated cigarettes (Multimedia Appendix 6). The participants (mean age 39.42, SD 11.03 years) were majority female (641/1160, 55.26%) and primarily self-identified as non-Hispanic White (667/1160, 57.5%); 19.22% (223/1160) self-identified as non-Hispanic Black; and 16.47% (191/1160) self-identified as of Hispanic ethnicity (Table 1). Over half (632/1160, 54.48%) of them reported annual household income less than US $50,000 and 39.57% (459/1160)

XSL•FO
RenderX

did not complete education beyond high school. On average, the participants reported smoking for 16.20 (SD 9.87) years; currently smoked 14.11 (SD 8.96) cigarettes per day and reported high levels of dependence on cigarettes (Tobacco Dependence Index-Cigarettes, mean 3.65, SD 0.81); 73.3% (850/1159) smoked mentholated cigarettes; and only 0.95% (11/1160) planned to quit smoking in the next 30 days at baseline (832/1160, 71.72% planned to never quit smoking).

Nearly three-quarters (825/1160, 71.12%) of the participants had ever used ENDS and, of those, 58.3% (481/825) had used ENDS in the 30 days preceding baseline (ie, were dual users). Dual users, on average, used ENDS 15.25 (SD 9.82) days out of the past 30 days. Among current ENDS users, the use of menthol-flavored products (183/481, 38%) was more common than use of tobacco-flavored ENDS (50/481, 10.4%). JUUL was selected as the primary ENDS brand by 34.5% (166/481) of those currently using ENDS.

XSL•FO
RenderX

Table 1. Sociodemographic and tobacco use characteristics of the overall sample (N=1160).

Sample characteristics	Values
JUUL2 flavor selection, n (%)	
Virginia Tobacco	242 (20.86)
Polar Menthol	239 (20.6)
Autumn Tobacco	219 (18.88)
Ruby Menthol	224 (19.31)
Summer Menthol	236 (20.34)
Sociodemographic characteristics	
Age (y), mean (SD)	39.42 (11.03)
Sex, n (%)	
Male	517 (44.57)
Female	641 (55.26)
Other	1 (0.09)
Prefer not to answer	1 (0.09)
Race or ethnicity, n (%)	
Hispanic	191 (16.47)
Non-Hispanic Black	223 (19.22)
Non-Hispanic other race[a]	66 (5.69)
Non-Hispanic White	667 (57.5)
Unknown	13 (1.12)
Marital status, n (%)	
Married	369 (31.81)
Living with partner	222 (19.14)
Divorced, separated, or widowed	167 (14.4)
Never married	379 (32.67)
Prefer not to say	23 (1.98)
Annual household income, n (%)	
<US $50,000	632 (54.48)
US $50,000-US $99,999	408 (35.17)
≥US $100,000	120 (10.34)
Highest level of education, n (%)	
High school graduate or less	459 (39.57)
Some college or trade school	422 (36.38)
College graduate or more education	279 (24.05)
Employment status, n (%)	
Full-time	754 (65)
Part-time	145 (12.5)
Other	261 (22.5)
Census region, n (%)	
Northeast	54 (4.66)
Midwest	356 (30.69)
South	606 (52.24)
West	144 (12.41)

XSL•FO
RenderX

Sample characteristics	Values
Cigarette smoking characteristics	
Smoke mentholated cigarettes, n (%)	850 (73.34)
Number of cigarettes smoked per smoking day, mean (SD)	14.11 (8.96)
Duration of smoking (y), mean (SD)	16.20 (9.87)
Age started smoking (y), mean (SD)	19.33 (6.27)
Cigarette dependence[b], mean (SD)	3.65 (0.81)
Plan to quit smoking in next 30 days, n (%)	11 (0.94)
Ever plan to quit smoking, n (%)	328 (28.28)
ENDS[c] use characteristics	
Ever used ENDS, n (%)	825 (71.12)
Age first used ENDS (y), mean (SD)	31.31 (11.4)
Ever used ENDS fairly regularly, n (%)	529 (64.12)
Used ENDS in past 30 days, n (%)	481 (58.3)
Number of days used ENDS in past 30 days, mean (SD)	15.25 (9.82)
Number of times used ENDS per use day, median (IQR)	9 (5-18)
ENDS dependence[b], mean (SD)	3.03 (0.97)
Primary ENDS flavor[d], n (%)	
Tobacco	50 (10.4)
Menthol	183 (38.1)
Mint	55 (11.4)
Fruit	152 (31.6)
Dessert or candy	33 (6.9)
Spice or clove	2 (0.4)
Some other flavor	6 (1.3)
Primary ENDS device type[d], n (%)	
Pod based	168 (34.9)
Disposable	229 (47.6)
Tank	56 (11.6)
Mod	8 (1.7)
Primary ENDS brand[d], n (%)	
JUUL	166 (34.5)
Vuse	63 (13.1)
Blu	51 (10.6)
NJOY	29 (6)
Puff bar	67 (13.9)
Other	105 (21.8)

[a]Non-Hispanic other race includes Asian, Pacific Islander or Native Hawaiian, American Indian or Alaska Native, and “another race not listed.”
[b]Tobacco Dependence Index in Population Assessment of Tobacco and Health adult survey (range 1-5; higher scores indicate greater dependence).
[c]ENDS: electronic nicotine delivery systems. Denominator is 481 past-30-day ENDS users with the exception of ever ENDS use.
[d]Participants selected the single flavor, nicotine concentration, or ENDS device they used most often.

XSL•FO
RenderX

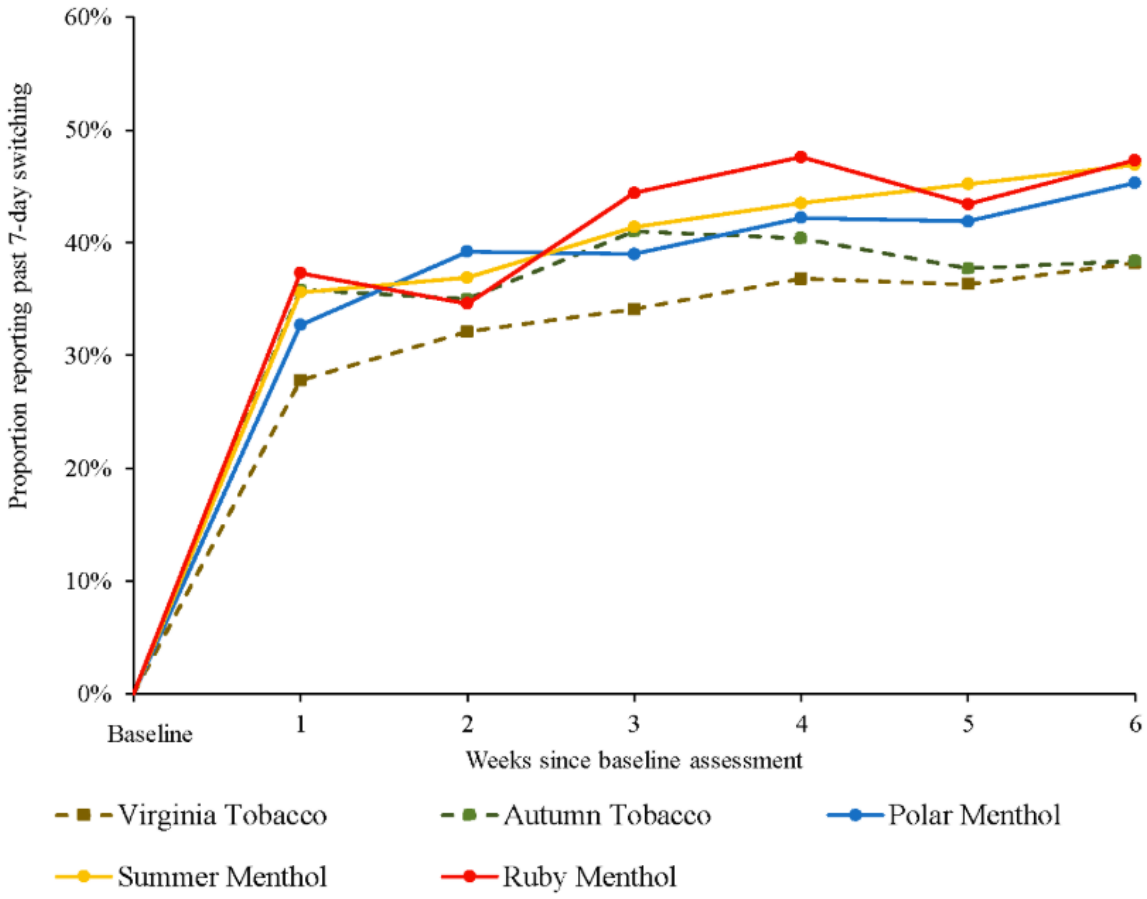
Analysis of Baseline Characteristics by Response to Weekly Surveys

In each flavor group, over 85% of participants who were enrolled and started the 6-week actual use period completed the week 6 survey (Virginia Tobacco, 207/242, 85.5%; Polar Menthol, 212/239, 88.7%; Autumn Tobacco, 190/219, 86.8%; Summer Menthol, 211/236, 89.4%; Ruby Menthol, 201/224, 89.7%) and over 70% completed all 6 weekly surveys (Virginia Tobacco, 178/242, 73.6%; Polar Menthol, 173/239, 72.4%; Autumn Tobacco, 158/219, 72.1%; Summer Menthol, 171/236, 72.5%; Ruby Menthol, 157/224, 70.1%). In the overall sample, there were a few statistically significant differences between the participants who completed all 6 weekly surveys and those who missed some surveys, and the magnitude of differences were small (Multimedia Appendix 7). Participants who completed all 6 surveys (vs 1-5 surveys) were significantly older, by 1.5 years (39.84 vs 38.34 years; *P*=.04), more likely to be female (481/837, 57.5% vs 160/323, 49.5%; *P*=.01), less likely to smoke menthol cigarettes (599/836, 71.7% vs 251/323, 77.7%; *P*=.04) and less likely to have ever used ENDS (575/837, 68.7% vs 250/323, 77.4%; *P*=.003). Among participants who had used ENDS, those who completed all 6 surveys initiated ENDS use about 2 years later (age 31.87 vs 30.01 years; *P*=.03). There were no significant differences by race or ethnicity, education, income, employment, plans to quit smoking, use of JUUL ENDS, or ENDS dependence among those using ENDS. Importantly, there were no differences in the heaviness of smoking or in cigarette dependence, which have been shown to be important predictors of switching. There were no significant differences in rates of survey completion among the JUUL2 flavor groups.

Switching Across 6-Week Actual Use Period

As displayed in Figure 1, in each of the 5 JUUL2 flavor groups, rates of past-7-day switching away from cigarettes increased across the 6-week actual use period: from 27.8% (62/223) at week 1 to 38.2% (79/207) at week 6 for Virginia Tobacco, 35.8% (73/204) at week 1 to 38.4% (73/190) at week 6 for Autumn Tobacco, 32.7% (73/223) at week 1 to 45.3% (96/212) at week 6 for Polar Menthol, 35.6% (77/216) at week 1 to 46.9% (99/211) at week 6 for Summer Menthol, and 37.3% (76/204) at week 1 to 47.3% (95/201) at week 6 for Ruby Menthol. In the combined sample that included all 5 flavors, there was a statistically significant linear increase in the rate of past-7-day switching over the 6-week actual use period (odds ratio [OR] 1.09, 95% CI 1.07-1.12; Multimedia Appendix 8).

Figure 1. Rates of past 7-day switching away from cigarettes across 6-week actual use period by JUUL2 flavor. Virginia Tobacco: week 1: n=223, week 2: n=234, week 3: n=223, week 4: n=220, week 5: n=215, week 6: n=207; Autumn Tobacco: week 1: n=204, week 2: n=206, week 3: n=200, week 4: n=203, week 5: n=191, week 6: n=190; Polar Menthol: week 1: n=223, week 2: n=222, week 3: n=213, week 4: n=218, week 5: n=210, week 6: n=212; Summer Menthol: week 1: n=216, week 2: n=217, week 3: n=210, week 4: n=214, week 5: n=210, week 6: n=211; Ruby Menthol: week 1: n=204, week 2: n=211, week 3: n=205, week 4: n=206, week 5: n=196, week 6: n=201.



XSL•FO
RenderX

At week 6, rates of past-30-day switching were 24.3% (55/226) for Virginia Tobacco, 28.6% (58/203) for Autumn Tobacco, 31.4% (70/223) for Polar Menthol, 33% (69/209) for Ruby Menthol, and 33.9% (74/218) for Summer Menthol (Table 2).

Sensitivity analyses in which smoking (ie, not being switched) was imputed to missing observations showed very similar results, although the absolute imputed switch rates were slightly lower (Multimedia Appendix 9). At week 6, the rates of past-7-day switching using the intent-to-treat approach ranged from 32.6% (79/242; Virginia Tobacco group) to 42.4% (95/224; Ruby Menthol group) and the rates of past-30-day switching ranged from 22.7% (55/242, Virginia Tobacco group) to 31.4% (74/236, Summer Menthol group).

Table 2. Rates of past-30-day switching at week 6 by JUUL2 product flavor.

JUUL2 product flavor	Proportion reporting past-30-day switching, % (SE; 95% CI)
Virginia Tobacco (N=226)	24.3 (2.9; 18.7-29.9)
Autumn Tobacco (N=203)	28.6 (3.2; 22.4-34.8)
Polar Menthol (N=223)	31.4 (3.1; 25.3-37.5)
Summer Menthol (N=218)	33.9[a] (3.2; 27.7-40.2)
Ruby Menthol (N=209)	33[a] (3.3; 26.6-39.4)

[a]Switch rate is significantly higher than Virginia Tobacco group (P<.05).

Association of JUUL2 Flavor Group and Menthol Cigarette Smoking With Switching

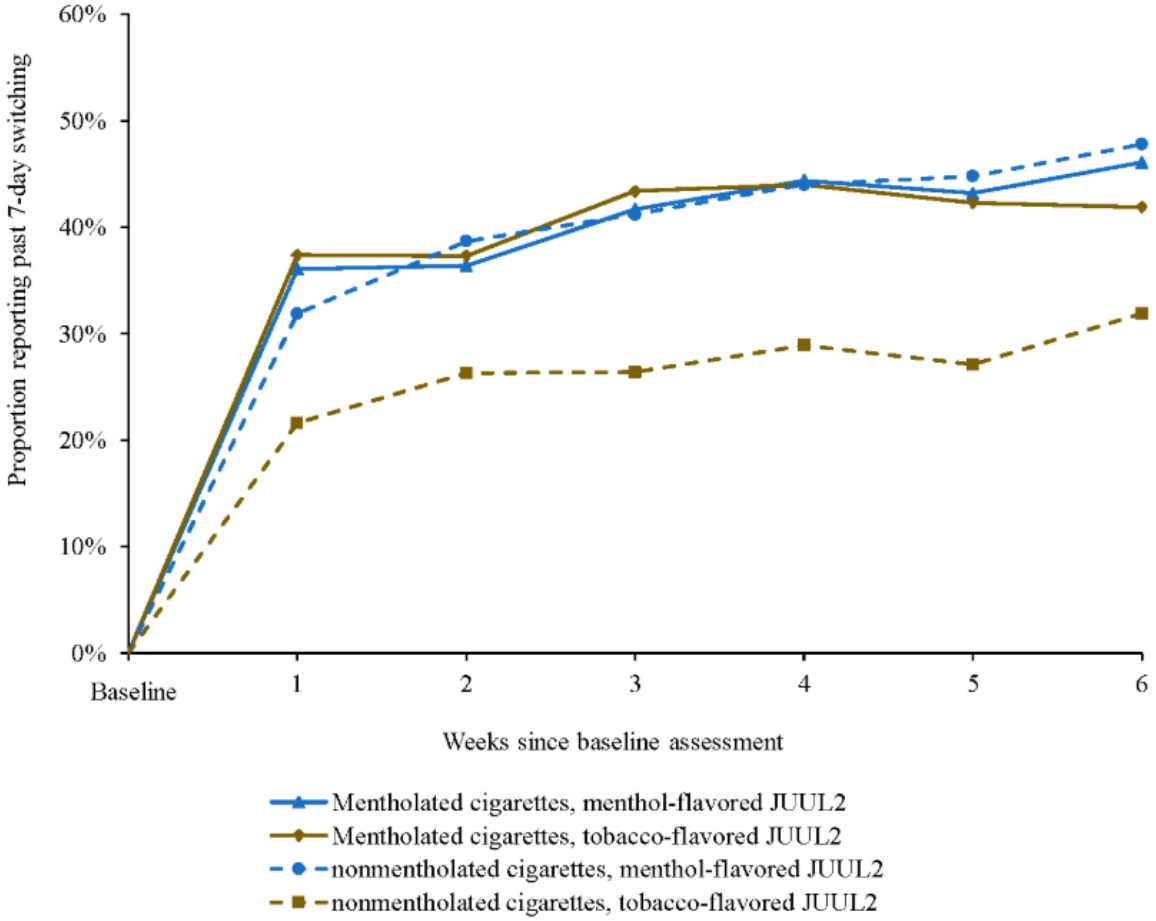
In the model assessing past-30-day switching at week 6, participants who used menthol-flavored (vs tobacco-flavored) JUUL2 products had statistically significant higher switch rates (OR 1.36, 95% CI 1.04-1.78; Multimedia Appendix 8)—there was no significant main effect of cigarette flavor (P=.06) and the interaction of JUUL2 flavor group and cigarette flavor was not significant (P=.10). Compared to the Virginia Tobacco group, rates of past-30-day switching were significantly higher in the Ruby Menthol group (OR 1.53, 95% CI 1.01-2.33) and Summer Menthol group (OR 1.60, 95% CI 1.06-2.42), respectively. These associations remained significant in the models that imputed smoking for missing data: Ruby Menthol versus Virginia Tobacco (OR 1.51, 95% CI 1.00-2.29) and Summer Menthol versus Virginia Tobacco (OR 1.55, 95% CI 1.03-2.33). The rate of past-30-day switching was numerically higher among the participants in the Polar Menthol group compared to the Virginia Tobacco group, but it did not show a statistically significant difference (OR 1.42, 95% CI 0.94-2.15).

The association of JUUL2 product flavor group and the rate of past-7-day switching was not significant as a main effect (OR 1.23, 95% CI 0.92-1.64; Multimedia Appendix 8). However, this association was significantly moderated by preferred cigarette flavor: among participants who smoked nonmentholated cigarettes, those who used menthol-flavored (vs tobacco-flavored) JUUL2 products had significantly higher switch rates (340/824, 41.3% vs 237/880, 26.9%; OR 1.71, 95% CI 1.04-2.83; Figure 2), whereas there was no significant difference in switch rates by JUUL2 product flavor among participants who smoked menthol cigarettes (1228/2975, 41.3% vs 668/1630, 41%; OR 0.88, 95% CI 0.65-1.18).

There was also a significant main effect of menthol cigarette smoking, with smokers of mentholated (vs nonmentholated) cigarettes demonstrating higher switch rates (OR 1.52, 95% CI 1.14-2.04; Multimedia Appendix 8), largely because of lower switch rates among participants who smoked nonmentholated cigarettes and used tobacco-flavored JUUL2 products (Figure 2). Among participants who smoked mentholated cigarettes, there were no statistically significant differences in sociodemographic or tobacco product use characteristics between those who selected tobacco-flavored and those who selected menthol-flavored JUUL2 products (Multimedia Appendix 10).

XSL•FO
RenderX

Figure 2. Past 7-day switching by JUUL2 product flavor and cigarette flavor across 6-week actual use period. Smoked menthol cigarettes, menthol-flavored JUUL2: week 1: n=499, week 2: n=508, week 3: n=492, week 4: n=504, week 5: n=482, week 6: n=490; smoked nonmentholated cigarettes, menthol-flavored JUUL2: week 1: n=144, week 2: n=142, week 3: n=136, week 4: n=134, week 5: n=134, week 6: n=134; smoked menthol cigarettes, tobacco-flavored JUUL2: week 1: n=273, week 2: n=287, week 3: n=274, week 4: n=273, week 5: n=265, week 6: N=258; smoked nonmentholated cigarettes, tobacco-flavored JUUL2: week 1: n=153, week 2: n=152, week 3: n=148, week 4: n=149, week 5: n=140, week 6: n=138.

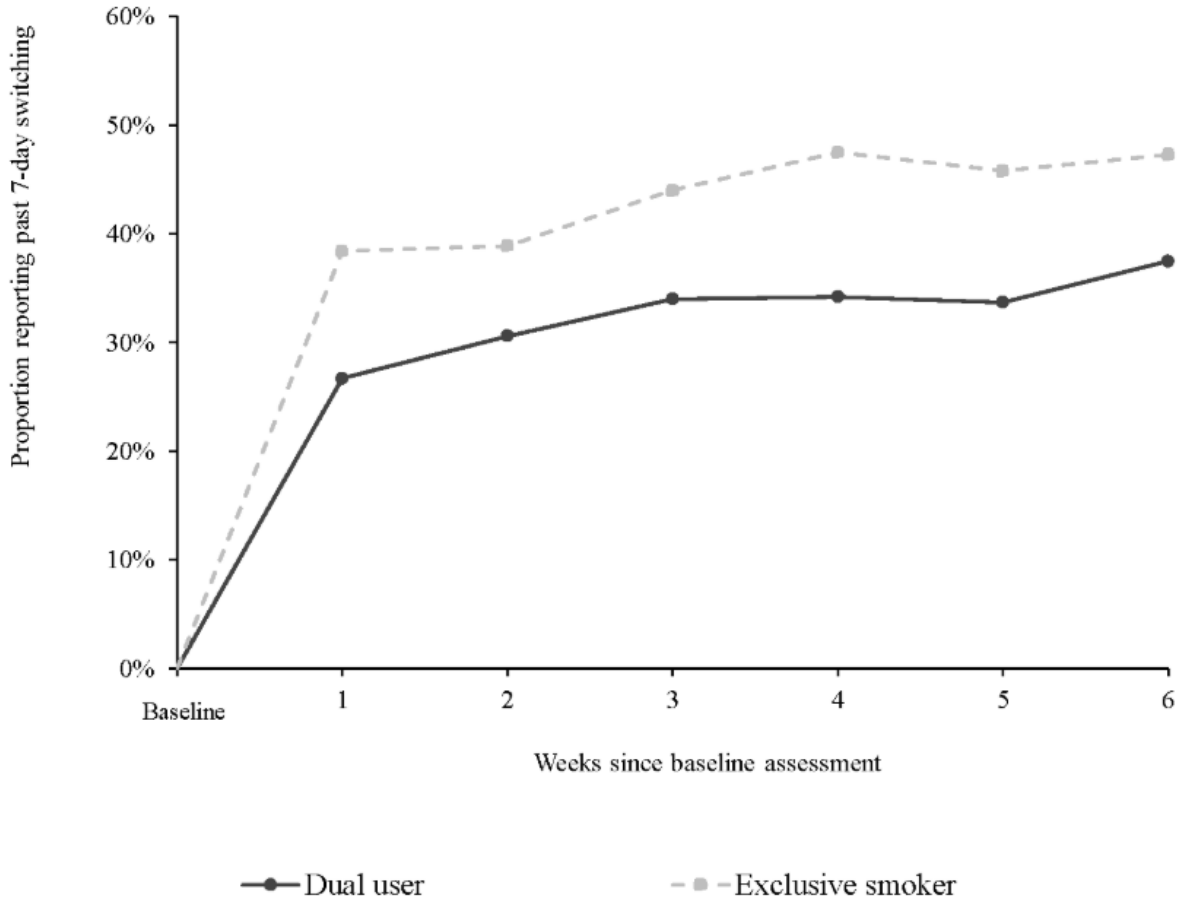


Differences in Switching by ENDS Use at Baseline

There were significant differences in the rates of past-7-day switching according to ENDS use in the 30 days preceding baseline: participants who were smoking but did not use ENDS in the past 30 days had significantly higher switch rates than the dual users (1630/3740, 43.6% vs 843/2575, 32.7%) across the 6-week actual use period (Figure 3 and Multimedia Appendix 11).

XSL•FO
RenderX

Figure 3. Past 7-days switching among baseline dual users and exclusive smokers across 6-week actual use period. Dual users: week 1: n=430, week 2: n=448, week 3: n=435, week 4: n=433, week 5: n=413, week 6: n=416; exclusive smokers: week 1: n=640, week 2: n=642, week 3: n=616, week 4: n=628, week 5: n=609, week 6: n=605.



Changes in Dependence and Respiratory Symptoms Among Participants Who Completely Switched (Did Not Smoke) in the Past 30 Days at Week 6

Among participants who reported past-30-day switching at week 6, levels of dependence on JUUL2 products were significantly lower than their own levels of dependence on combustible cigarettes at baseline, for all 5 flavors (P<.001; Figure 4).

Participants who did not smoke in the past 30 days at week 6 also experienced significant decreases in frequency of self-reported respiratory symptoms relative to baseline when they were smoking cigarettes, across all 5 flavors (P≤.01; Figure 5).

XSL•FO
RenderX

Figure 4. Change from baseline cigarette dependence to JUUL2 dependence at week 6 among the past-30-day switchers (mean, SE). *Statistically significant difference between levels of dependence on cigarettes at baseline and levels of dependence on JUUL2 products at week 6 (*P*<.001).

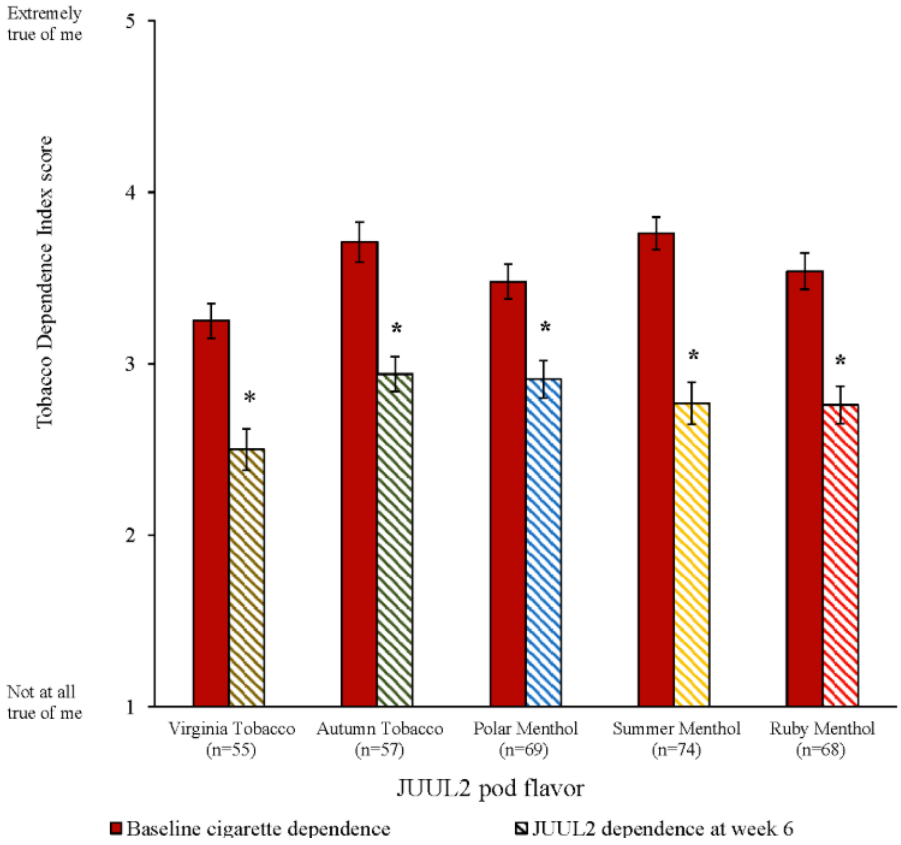
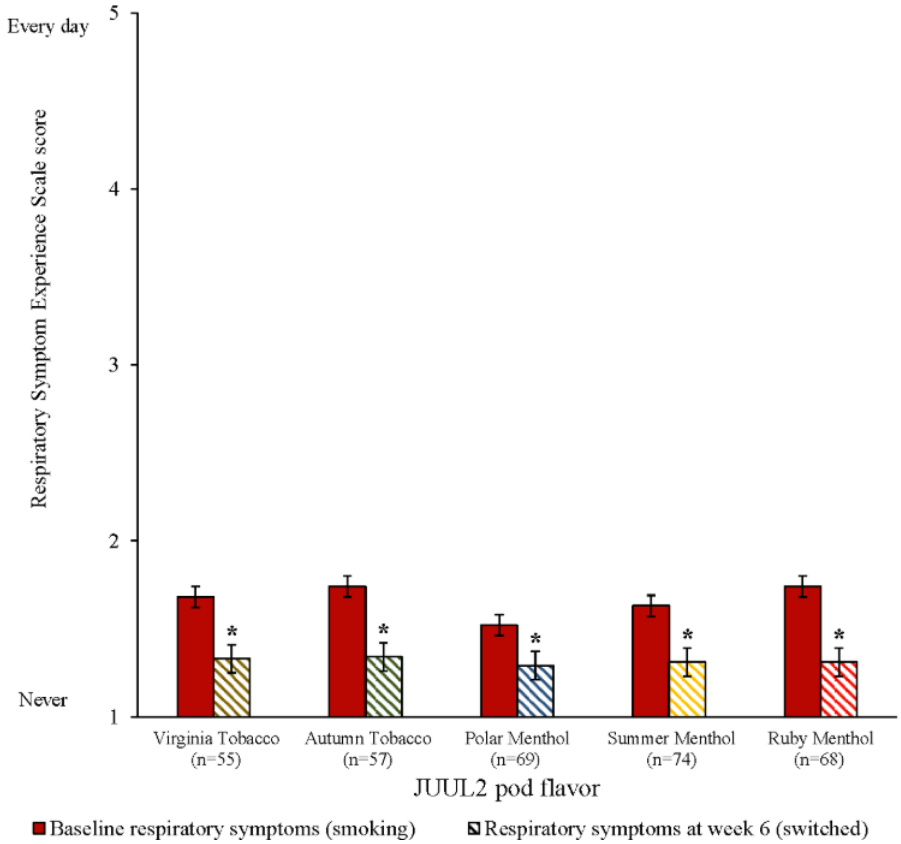


Figure 5. Change in frequency of respiratory symptoms from baseline to week 6 among the past-30-day switchers (mean, SE). *Statistically significant difference between frequency of respiratory symptoms at baseline and at follow-up (P≤.01).

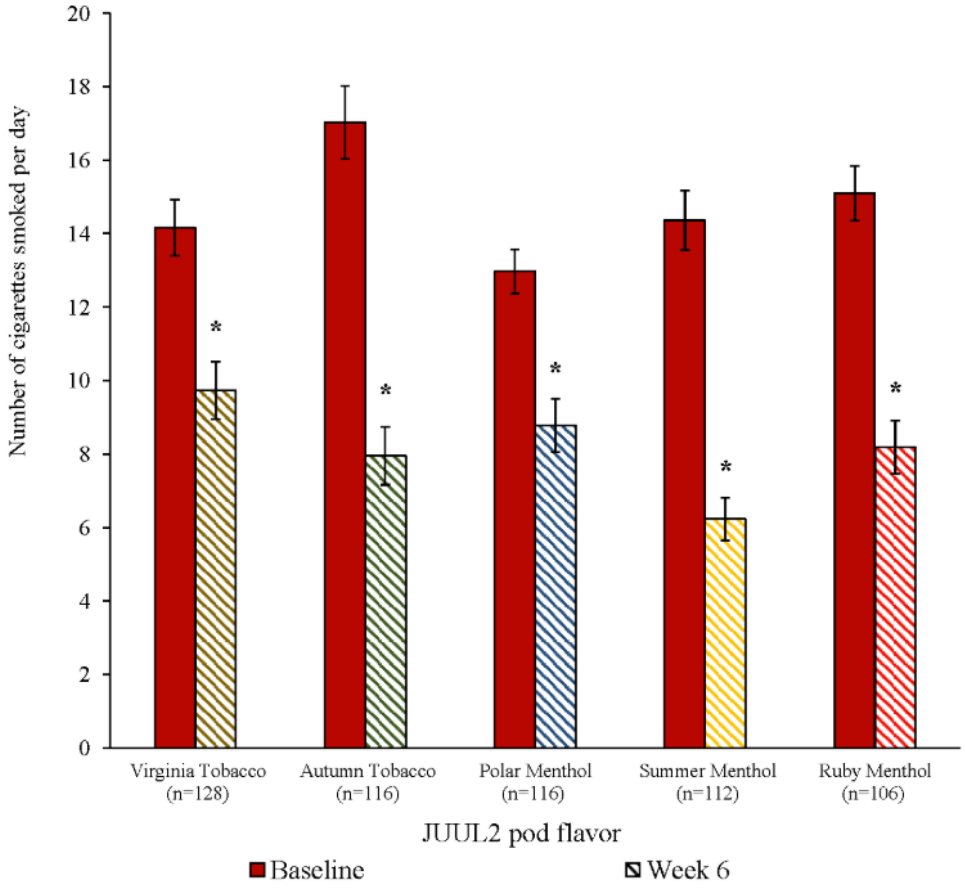


XSL•FO
RenderX

Smoking Reduction Among Participants Who Continued Smoking at Week 6

Among participants who continued to smoke at week 6, average daily cigarette consumption in all 5 JUUL2 flavor groups was significantly reduced relative to participants' own cigarette consumption at baseline (P<.001; Figure 6); across all flavor groups, the proportion of participants that reported reducing their daily cigarette consumption by at least 50% at week 6 relative to baseline ranged from 50.9% (59/116) to 62.9% (73/116; Multimedia Appendix 12).

Figure 6. Change in daily cigarette consumption from baseline to week 6 among participants who did not switch (mean, SE). *Statistically significant difference in number of cigarettes per day smoked at baseline and week 6 among participants who continued smoking (P<.001).



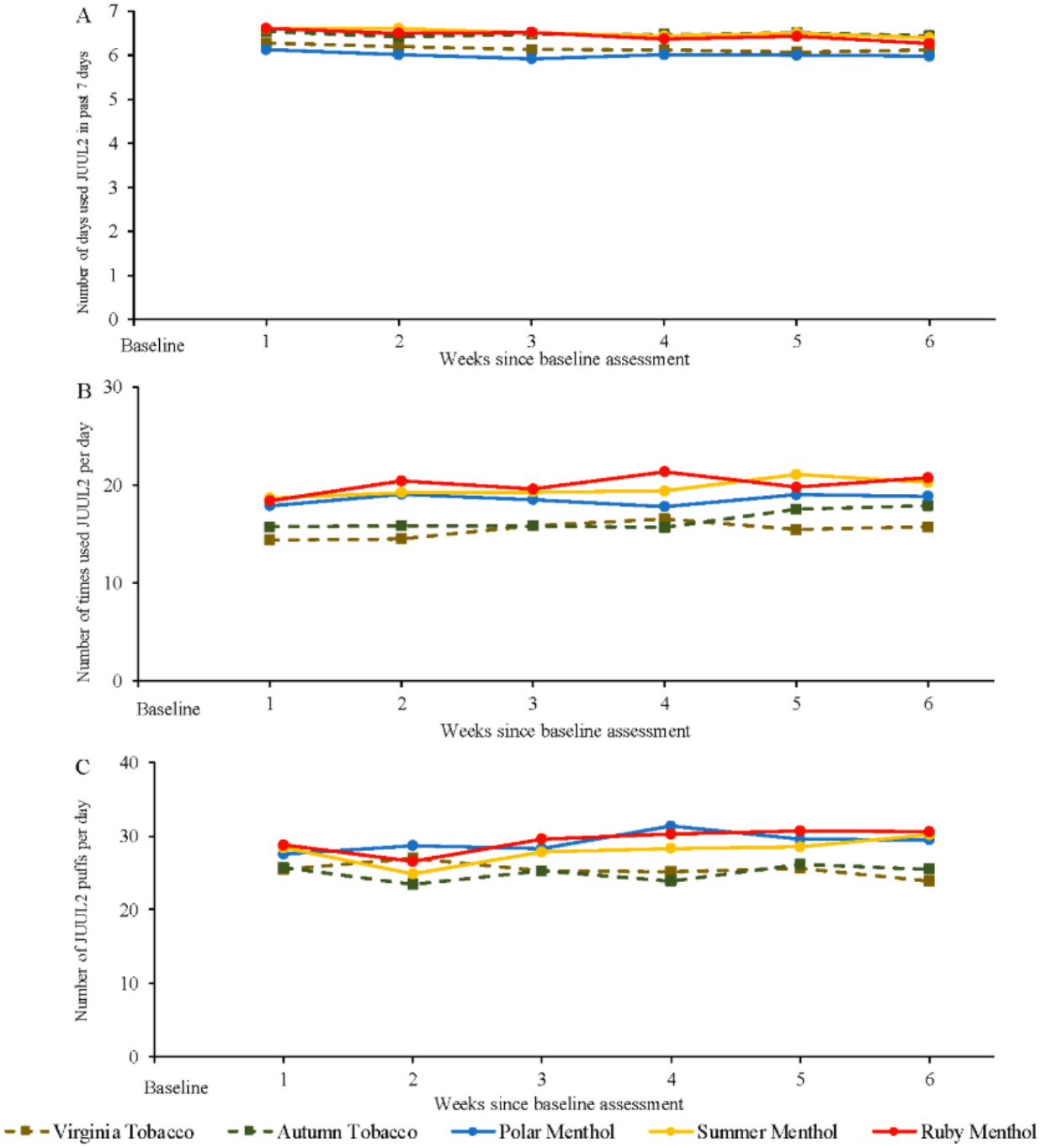
Patterns of JUUL2 Product Use and Subjective Effects

Across the 6-week actual use period, the prevalence of past-7-day JUUL2 product use was over 97% in each of 5 flavor groups (Multimedia Appendix 13). In each of the 5 flavor groups, the median frequency of JUUL2 product use at each weekly survey was 7 days of use in the past 7 days (ie, daily use; Figure 7A); participants generally reported using JUUL2 products between 15 and 20 times per day (Figure 7B) and 25 to 30 puffs per day (Figure 7C). Across the 6-week period, there were no statistically significant linear increases in JUUL2 use occasions per day in the Virginia Tobacco and Polar Menthol JUUL2 groups (P>.15; Multimedia Appendix 14); there were statistically significant linear increases over time in use occasions per day among participants who used Autumn Tobacco, Summer Menthol, and Ruby Menthol JUUL2 products but the magnitude of the increases was small (unstandardized β coefficients=.45−.54; Multimedia Appendix 14). Across the 6-week period, there were no statistically significant increases in JUUL2 puffs per day in the Virginia Tobacco and Autumn Tobacco JUUL2 groups (P>.09); there were statistically significant linear increases over time in puffs per day among participants who used Polar Menthol, Summer Menthol, and Ruby Menthol JUUL2 products but the magnitude of the increases was small (unstandardized β coefficients=.43−.60; Multimedia Appendix 14).

Subjective responses to the 5 JUUL2 products were similar across 5 flavors groups at the week 6 survey: ratings of satisfaction approximated "A lot," ratings of psychological reward approximated "Moderately," ratings of aversion approximated "Very little," and ratings of relief approximated "A lot" (Multimedia Appendix 15). When combined across all 5 flavors, ratings of subjective satisfaction (ie, modified Product Evaluation Scale Satisfaction subscale) at week 2 were significantly associated with higher likelihood of complete past-30-day switching at week 6 (OR 1.25, 95% CI 1.11-1.40).

XSL•FO
RenderX

Figure 7. Patterns of JUUL2 product use across 6-week actual use period. (A) Number of days used JUUL2 in past 7 days; (B) number of times used JUUL2 per day; and (C) number of JUUL2 puffs per day.



Adverse Events

The incidence of study-emergent AEs was low for each of the 5 JUUL2 products during the 6-week actual use period—in each group less than 2.5% of participants reported an AE (Virginia Tobacco, 1/262, 0.4%; for, Autumn Tobacco, 2/237, 0.8%, Polar Menthol, 2/262, 0.8%, Ruby Menthol 2/234, 0.9%, for Summer Menthol, 6/249, 2.4%; Multimedia Appendices 16 and 17). No serious AEs were reported in the study and all evaluable AEs were reported as mild or moderate in intensity and resolved during the study.

Discussion

Principal Findings

This actual use study enrolled a sample of >1000 US adults who smoked cigarettes daily, following them for 6 weeks as they used JUUL2 ENDS with their preferred flavor in their natural environment. The sample was sociodemographically diverse and predominantly included US adults who smoked mentholated cigarettes, were not ready to quit smoking (ie, were not planning to quit within 30 days), and had experience using ENDS. On average, participants were highly dependent on cigarettes when entering the study. The results demonstrate that the use of JUUL2 ENDS products was associated with high rates of

XSL•FO
RenderX

completely switching away from cigarettes among US adults who smoke cigarettes, including both those who smoke mentholated cigarettes and nonmentholated cigarettes and users and nonusers of ENDS. Moreover, participants who reported past-30-day switching at the end of the study experienced meaningful reductions in dependence on tobacco products and improvements in respiratory symptoms during this 6-week period.

Despite a profile that did not favor smoking cessation (eg, high levels of cigarette dependence, almost no participants intending to quit smoking in 30 days, and many never planning to quit), rates of 7-day point prevalence abstinence exceeded 40% after 6 weeks, and at week 6 over 24% of participants in each of the 5 flavor groups (Virginia Tobacco, 55/226, 24.3%; Autumn Tobacco, 58/203, 28.6%, Polar Menthol, 70/223, 31.4%, Ruby Menthol, 69/209, 33%, Summer Menthol, 74/218, 33.9%) reported they had not smoked in the previous 30 days—a criterion often used to evaluate smoking cessation in clinical trials [25]. The proportion reporting switching away from smoking increased over the 6-week period, a pattern that was observed to continue over longer periods in previous observational and randomized studies of JUUL product users [8,26]. Furthermore, the high rates of switching are consistent with participants' reports of high levels of subjective satisfaction in this study and in experimental studies evaluating JUUL2 products [17], which is an established predictor of switching behavior [26] that was confirmed in this study.

It was notable that high switching rates were observed even though participants reported relatively light use of JUUL2, reportedly averaging around 25 to 30 puffs per day. One factor for this observation may be that the participants appear to have spread their use over time, engaging in "grazing" [27]—they reported averaging 15 to 20 use occasions per day, implying just 1 or 2 puffs per occasion. The volume of aerosol produced by the JUUL2 device may also have made puffs more impactful, mitigating a need for more puffs. In any case, more research on patterns of use associated with switching, including using objective measures of puff topography, would be useful.

Participants who used menthol-flavored (vs tobacco-flavored) JUUL2 products had significantly higher rates of past-30-day switching at the end of the 6-week actual use period. In addition, the association of JUUL2 flavor group and past-7-day switching significantly varied by menthol cigarette smoking (ie, the JUUL2 flavor and menthol cigarette interaction term was statistically significant): participants who smoked nonmentholated cigarettes and used menthol-flavored (vs tobacco-flavored) JUUL2 products demonstrated significantly higher switch rates, but this was not true for those who smoked mentholated cigarettes. The finding that participants who smoked nonmentholated cigarettes and used menthol-flavored ENDS had higher switch rates than those using tobacco-flavored ENDS is consistent with observational studies of US adults who smoke and use ENDS, including JUUL products [28,29], and an actual use study that evaluated the use of heated tobacco products [15]. This suggests that the availability of menthol-flavored ENDS may particularly benefit adults who smoke nonmentholated cigarettes.

Participants who smoked cigarettes but did not use ENDS when they entered the study had significantly higher rates of past-7-day switching compared to those who were dual users (ie, used ENDS and smoked cigarettes concurrently) across the 6-week actual use period. Ongoing dual use may be a behavioral marker for difficulty in stopping smoking with ENDS, as the dual users were, on average, relatively frequent ENDS users at baseline but had not yet completely switched. Nonetheless, both baseline dual users and baseline non-ENDS users displayed high rates of past-7-day switching at the end of the study.

Among participants who reported switching for the past 30 days at week 6, levels of dependence on JUUL2 products were statistically significant and meaningfully lower than participants' own levels of dependence on cigarettes when they entered the study; the decrease exceeded the minimally important difference for the measure identified in a prior observational study [30]. This finding concords with a large body of evidence indicating that ENDS, including JUUL products, produce lower levels of dependence than cigarettes [1,30-32]. In addition, consumption of JUUL2 products minimally increased across the 6-week actual use period (increases of less than one use occasion and puff per week). In accordance with data from controlled laboratory studies [17], these results demonstrate that JUUL2 products have lower abuse liability than cigarettes.

Participants who completely switched to JUUL2 products for at least 30 days at the completion of the study also reported significantly lower frequency of respiratory symptoms relative to when they entered the study and were smoking cigarettes. This improvement in respiratory symptoms was statistically significant but modest in magnitude—below the minimally important difference of 0.57 proposed in a prior psychometric validation study (operationalized as the difference in frequency of respiratory symptoms between participants with vs without respiratory symptom-relevant diagnoses, including chronic obstructive pulmonary disease) [21]. However, these improvements in respiratory symptoms occurred after relatively short periods of smoking abstinence and 6 weeks of JUUL2 product use, thus it seems likely that respiratory symptoms will continue to improve over longer periods of switching. For example, a study of adults who switched to JUUL products for an average of 3 years found that they had meaningfully lower respiratory symptoms than a matched comparison group that had continued smoking [33]. In this study, the fact that participants reported low baseline respiratory symptom scores when entering the study (mean <2; approximating ratings of "rarely [1-5 days]"), suggested that changes were limited by floor effects.

Substantial proportions of participants who persisted in smoking (ie, were dual users) after 6 weeks of being provided with JUUL2 products significantly reduced their daily cigarette consumption by at least 50% compared to when they entered the study. These levels of reductions in cigarette consumption have been shown to meaningfully reduce exposure to toxicants in cigarette smoke that are associated with smoking-related diseases [34-36] and have been suggested to reduce risk of smoking-related disease [37]. Finally, the JUUL2 products were well tolerated, producing only infrequent and minor AEs.

XSL•FO
RenderX

Strengths and Limitations

Strengths of the study include the large sociodemographically diverse sample of US adults who smoked every day and reported high levels of cigarette dependence, including large numbers of participants who were not planning to ever quit smoking and smoked mentholated cigarettes. Allowing participants to select their preferred flavor for use during the 6-week actual use period replicated real-world consumer product use and allowed attribution of switching behavior to a single JUUL2 product. However, participants were not allowed to change their selected flavor during the 6-week actual use period, as they would be able to in the real world. In addition, this self-selection of flavors and lack of randomization should be considered when interpreting differences in switch rates between the flavor groups.

Rates of retention were high throughout the 6-week actual use period. In addition, comparisons of sociodemographic and tobacco product characteristics among participants who completed all 6 surveys (vs those who missed some surveys) showed that participants who missed some surveys were not meaningfully different from those who completed all the surveys, and were not heavier or more dependent smokers who would have been expected to have lower rates of switching, suggesting low likelihood of bias due to nonresponse and missing observations. Analyses imputing smoking to all missing observations were performed, and yielded results that paralleled those based on observed data, indicating a lack of material bias in survey response.

Only individuals who reported sufficient levels of interest in using JUUL2 products were eligible to enroll in the study, which may have led to higher switch rates, but also enhanced ecological validity by mirroring real-world use of consumer products, where people buy and adopt products they are interested in, and are not arbitrarily prescribed or assigned products in which they have little interest. Assignment to individual flavors was based on expressed interest, which again is consistent with real-world use. In any case, most participants in both study arms expressed sufficient interest in all the flavors offered to them. Despite this interest in multiple flavors, most of the participants did show substantial interest in one or more flavors offered to them. This is important, as it suggests that making a range of flavors available is likely to increase the proportion of adults who smoke adopt ENDS, thus increasing reach of ENDS and potential switching away from smoking.

Participants were provided with JUUL2 study products at no charge, which is not consistent with the real world, and could have resulted in heavier product use and greater retention or influenced switching. The actual use portion of the study was only 6 weeks in duration and there was no follow-up assessment after completion. Future studies could evaluate trajectories of switching among adults who smoke and adopt JUUL2 products over longer periods. A large study of real-world US adult JUUL purchasers showed that switch rates continued to increase over time, reaching 58% after 2 years of initial purchase [6].

A limitation was the analyses' reliance on self-report; cigarette smoking and JUUL2 product use were not biochemically verified, and respiratory symptoms were evaluated with a self-report measure. Although this is standard practice for actual use studies [12-14] and observational population-based studies more broadly [24], biomarker data could be useful in future studies. In addition, since the primary aim was descriptive (ie, assessing the proportion of smokers who switched), and did not involve any inferential tests, no adjustment was made for multiplicity. Causal inferences regarding switching are limited by the lack of a control group that did not receive JUUL2 products, hence it is not possible to determine how many participants would have switched or stopped smoking if they were not provided with JUUL2 products. However, almost all participants in the sample did not plan to quit smoking in the next 30 days at baseline, and many were not planning to ever quit, suggesting that quit rates would have been very low.

Conclusions

The data from this 6-week actual use study demonstrate that adoption of JUUL2 products is associated with substantial rates of completely switching away from cigarettes among US adults who smoke, and meaningful reductions in cigarette consumption among adults who do not completely switch after 6 weeks. Participants who completely switched from smoking cigarettes to use of JUUL2 products at the end of 6 weeks experienced significant reductions in dependence and improvements in respiratory symptoms. Switch rates were higher among participants who used menthol-flavored (vs tobacco-flavored) JUUL2 products, particularly among adults who smoke nonmentholated cigarettes. Although this study primarily focused on behavioral end points, when considered with data from clinical biomarker studies, these changes in tobacco product use behaviors (ie, complete switching and substantial smoking reduction) are likely to decrease exposure to harmful cigarette-related toxicants, and to ultimately reduce risk of tobacco-related diseases.

Acknowledgments

This study was funded by Juul Labs, Inc. The authors would like to thank Lauren C Smith for her assistance with the conduct and implementation of the study.

Authors' Contributions

NIG led the conceptualization of the manuscript, wrote the manuscript text, and prepared the tables and figures. Authors MAS and SP conducted the analyses and assisted in preparation of the tables and figures. Authors RAB, SS, and AS aided in

XSL•FO
RenderX

conceptualization and writing of the manuscript and provided feedback on drafts. All authors contributed to and approved of the final manuscript.

Conflicts of Interest

NIG and RAB are full-time employees of Juul Labs, Inc. MAS, AS, and SP are full-time employees of Pinney Associates, Inc, and SS is a senior adviser to Pinney Associates, Inc. Pinney Associates has consulted Juul Labs, Inc on tobacco harm reduction since October 2019. As of October 2024, Pinney Associates also consults Philip Morris International solely on US regulatory pathways for noncombustible nontobacco nicotine products. Pinney Associates does not consult on combustible tobacco products. AS also individually provides consulting services to the Center of Excellence for the Acceleration of Harm Reduction through ECLAT Srl, which received funding from the Foundation for a Smoke-Free World.

Multimedia Appendix 1
A schematic diagram of the study design.
[PDF File (Adobe PDF File), 167 KB - ijmr_v14i1e60620_app1.pdf]

Multimedia Appendix 2
The study sites in traditional and complex flavor arms.
[PDF File (Adobe PDF File), 133 KB - ijmr_v14i1e60620_app2.pdf]

Multimedia Appendix 3
Participant flowchart—traditional flavors study arm.
[PDF File (Adobe PDF File), 143 KB - ijmr_v14i1e60620_app3.pdf]

Multimedia Appendix 4
Participant flowchart—complex flavors study arm.
[PDF File (Adobe PDF File), 145 KB - ijmr_v14i1e60620_app4.pdf]

Multimedia Appendix 5
Ratings of interest in JUUL2 products and flavors at the end of trial week.
[PDF File (Adobe PDF File), 102 KB - ijmr_v14i1e60620_app5.pdf]

Multimedia Appendix 6
Sociodemographic and tobacco use characteristics among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 196 KB - ijmr_v14i1e60620_app6.pdf]

Multimedia Appendix 7
Sociodemographic and tobacco use characteristics by number of assessments completed.
[PDF File (Adobe PDF File), 267 KB - ijmr_v14i1e60620_app7.pdf]

Multimedia Appendix 8
Association of JUUL2 flavor group and cigarette flavor with past 7-day and past 30-day switch rates across 6-week actual use period.
[PDF File (Adobe PDF File), 116 KB - ijmr_v14i1e60620_app8.pdf]

Multimedia Appendix 9
Rates of past 30-day and past 7-day switching away from cigarettes across 6-week actual use period by JUUL2 flavor imputing missing as smoking (intent-to-treat).
[PDF File (Adobe PDF File), 122 KB - ijmr_v14i1e60620_app9.pdf]

Multimedia Appendix 10
Sociodemographic and tobacco use characteristics of participants who smoked nonmentholated cigarettes by JUUL2 flavor selection.
[PDF File (Adobe PDF File), 149 KB - ijmr_v14i1e60620_app10.pdf]

Multimedia Appendix 11

XSL•FO
RenderX

Association of past 30-day electronic nicotine delivery systems use at baseline with past 7-day switch rates across 6-week actual use period.
[PDF File (Adobe PDF File), 129 KB - ijmr_v14i1e60620_app11.pdf]

Multimedia Appendix 12
Change in daily cigarette consumption from baseline to week 6 among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 149 KB - ijmr_v14i1e60620_app12.pdf]

Multimedia Appendix 13
Patterns of past 7-day JUUL2 product use across 6-week actual use period among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 178 KB - ijmr_v14i1e60620_app13.pdf]

Multimedia Appendix 14
Temporal trends in daily JUUL2 product use across 6-week actual use period among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 121 KB - ijmr_v14i1e60620_app14.pdf]

Multimedia Appendix 15
Subjective responses to JUUL2 products at week 6 survey among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 134 KB - ijmr_v14i1e60620_app15.pdf]

Multimedia Appendix 16
Adverse events—traditional and complex flavors trial week.
[PDF File (Adobe PDF File), 137 KB - ijmr_v14i1e60620_app16.pdf]

Multimedia Appendix 17
Adverse events over 6-week actual use period among JUUL2 flavor groups.
[PDF File (Adobe PDF File), 179 KB - ijmr_v14i1e60620_app17.pdf]

References

1. National Academies of Sciences, Engineering, and Medicine. Public Health Consequences of E-Cigarettes. Washington, DC: The National Academies Press; 2018.
2. Balfour DJ, Benowitz NL, Colby SM, Hatsukami DK, Lando HA, Leischow SJ, et al. Balancing consideration of the risks and benefits of e-cigarettes. Am J Public Health 2021 Sep;111(9):1661-1672. [doi: 10.2105/AJPH.2021.306416] [Medline: 34410826]
3. Gottlieb S, Zeller M. A nicotine-focused framework for public health. N Engl J Med 2017 Sep 21;377(12):1111-1114. [doi: 10.1056/NEJMp1707409] [Medline: 28813211]
4. Toll BA, Smith TT, King BA. Nicotine e-cigarettes: considerations for healthcare providers. Nat Med 2024 Jun;30(6):1513-1514. [doi: 10.1038/s41591-024-02926-7] [Medline: 38627561]
5. Kasza KA, Tang Z, Seo YS, Benson AF, Creamer MR, Edwards KC, et al. Divergence in cigarette discontinuation rates by use of electronic nicotine delivery systems (ENDS): longitudinal findings from the United States PATH study waves 1-6. Nicotine Tob Res 2025 Jan 22;27(2):236-243. [doi: 10.1093/ntr/ntae027] [Medline: 38566367]
6. Kim S, Goldenson NI, Selya A, Shiffman S. Switching away from smoking and reduction in cigarette consumption among U.S. adult purchasers of the JUUL system across 24 months including diverse subpopulations disproportionately affected by cigarette smoking. Nicotine Tob Res 2024 Aug 22;26(9):1183-1191. [doi: 10.1093/ntr/ntae072] [Medline: 38553983]
7. Pulvers K, Nollen NL, Rice M, Schmid CH, Qu K, Benowitz NL, et al. Effect of pod e-cigarettes vs cigarettes on carcinogen exposure among African American and Latinx smokers: a randomized clinical trial. JAMA Netw Open 2020 Nov 02;3(11):e2026324 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.26324] [Medline: 33206193]
8. Xu Y, Goldenson NI, Prakash S, Augustson EM, Shiffman S. Randomized trial assessing the effect of the JUUL system on switching away from cigarettes and smoking reduction among U.S. adults who smoke cigarettes. Exp Clin Psychopharmacol 2024 Feb;32(1):3-15. [doi: 10.1037/pha0000698] [Medline: 38127516]
9. Carpenter MJ, Wahlquist AE, Dahne J, Gray KM, Cummings KM, Warren G, et al. Effect of unguided e-cigarette provision on uptake, use, and smoking cessation among adults who smoke in the USA: a naturalistic, randomised, controlled clinical trial. EClinicalMedicine 2023 Sep;63:102142 [FREE Full text] [doi: 10.1016/j.eclinm.2023.102142] [Medline: 37753443]
10. Bradford DC, McCamon DL, Page BC, Richardson CA, Hunter KA. The actual use trial: a description of design principles and methods. SelfCare 2010;1(4):117-123 [FREE Full text]

XSL•FO
RenderX

11. Csoke E, Landes S, Francis MJ, Ma L, Teotico Pohlhaus D, Anquez-Traxler C. How can real-world evidence aid decision making during the life cycle of nonprescription medicines? Clin Transl Sci 2022 Jan 15;15(1):43-54 [FREE Full text] [doi: 10.1111/cts.13129] [Medline: 34405554]
12. Campbell C, Feehan M, Kanitscheider C, Makena PS, Cai J, Baxter SA. Designing studies to inform tobacco harm reduction: learnings from an oral nicotine pouch actual use pilot study. JMIR Form Res 2022 Aug 19;6(8):e37573 [FREE Full text] [doi: 10.2196/37573] [Medline: 35984682]
13. Becker E, McCaffrey S, Lewis J, Vansickel A, Larson E, Sarkar M. Characterization of Ad Libitum use behavior of On! Nicotine pouches. Am J Health Behav 2023 Jun 30;47(3):428-449. [doi: 10.5993/AJHB.47.3.1] [Medline: 37596760]
14. Roulet S, Chrea C, Kanitscheider C, Kallischnigg G, Magnani P, Weitkunat R. Potential predictors of adoption of the Tobacco Heating System by U.S. adult smokers: an actual use study. F1000Res 2019 Jul 14;8:214 [FREE Full text] [doi: 10.12688/f1000research.17606.2] [Medline: 31559015]
15. Karelitz JL, He Y, Becker E, Vansickel A. Switching behavior and changes in smoking behavior by menthol cigarette preference and menthol heated tobacco product use among adults in the United States who smoke cigarettes: an actual use study. Harm Reduct J 2025 Feb 20;22(1):19 [FREE Full text] [doi: 10.1186/s12954-025-01170-7] [Medline: 39980000]
16. Goldenson NI, Augustson EM, Chen J, Shiffman S. Pharmacokinetic and subjective assessment of prototype JUUL2 electronic nicotine delivery system in two nicotine concentrations, JUUL system, IQOS, and combustible cigarette. Psychopharmacology (Berl) 2022 Mar 20;239(3):977-988 [FREE Full text] [doi: 10.1007/s00213-022-06100-0] [Medline: 35184228]
17. Goldenson NI, Shiffman S, Sembower MA, Black RA. Assessment of abuse liability and switching potential of menthol-flavored pod-based electronic nicotine delivery systems among US adults who smoke cigarettes. Drug Alcohol Depend 2024 May 01;258:111279 [FREE Full text] [doi: 10.1016/j.drugalcdep.2024.111279] [Medline: 38614020]
18. Cook DK, Lalonde G, Oldham MJ, Wang J, Bates A, Ullah S, et al. A practical framework for novel electronic nicotine delivery system evaluation: chemical and toxicological characterization of JUUL2 aerosol and comparison with reference cigarettes. Toxics 2024 Jan 04;12(1):41 [FREE Full text] [doi: 10.3390/toxics12010041] [Medline: 38250996]
19. Strong DR, Leas E, Noble M, White M, Frissell KC, Glasser A, et al. Predictive validity of the adult tobacco dependence index: findings from waves 1 and 2 of the Population Assessment of Tobacco and Health (PATH) study. Drug Alcohol Depend 2020 Sep 01;214:108134 [FREE Full text] [doi: 10.1016/j.drugalcdep.2020.108134] [Medline: 32629146]
20. Strong DR, Pearson J, Ehlke S, Kirchner T, Abrams D, Taylor K, et al. Indicators of dependence for different types of tobacco product users: descriptive findings from Wave 1 (2013-2014) of the Population Assessment of Tobacco and Health (PATH) study. Drug Alcohol Depend 2017 Sep 01;178:257-266 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.05.010] [Medline: 28675817]
21. Shiffman S, McCaffrey SA, Hannon MJ, Goldenson NI, Black RA. A new questionnaire to assess respiratory symptoms (the respiratory symptom experience scale): quantitative psychometric assessment and validation study. JMIR Form Res 2023 Apr 14;7:e44036 [FREE Full text] [doi: 10.2196/44036] [Medline: 37058347]
22. Hatsukami DK, Zhang Y, O'Connor RJ, Severson HH. Subjective responses to oral tobacco products: scale validation. Nicotine Tob Res 2013 Jul;15(7):1259-1264 [FREE Full text] [doi: 10.1093/ntr/nts265] [Medline: 23239843]
23. Clopper CJ, Pearson ES. The use of confidence or fiducial limits illustrated in the case of the binomial. Biometrika 1934 Dec;26(4):404-413. [doi: 10.2307/2331986]
24. Blankers M, Smit ES, van der Pol P, de Vries H, Hoving C, van Laar M. The missing=smoking assumption: a fallacy in internet-based smoking cessation trials? Nicotine Tob Res 2016 Jan;18(1):25-33. [doi: 10.1093/ntr/ntv055] [Medline: 25744969]
25. Piper ME, Bullen C, Krishnan-Sarin S, Rigotti NA, Steinberg ML, Streck JM, et al. Defining and measuring abstinence in clinical trials of smoking cessation interventions: an updated review. Nicotine Tob Res 2020 Jun 12;22(7):1098-1106 [FREE Full text] [doi: 10.1093/ntr/ntz110] [Medline: 31271211]
26. Goldenson NI, Shiffman S, Hatcher C, Lamichhane D, Gaggar A, Le GM, et al. Switching away from cigarettes across 12 months among adult smokers purchasing the JUUL system. Am J Health Behav 2021 May 01;45(3):443-463. [doi: 10.5993/ajhb.45.3.4]
27. Notley C, Ward E, Dawkins L, Holland R. The unique contribution of e-cigarettes for tobacco harm reduction in supporting smoking relapse prevention. Harm Reduct J 2018 Jun 20;15(1):31 [FREE Full text] [doi: 10.1186/s12954-018-0237-7] [Medline: 29921278]
28. Mok Y, Jeon J, Levy DT, Meza R. Associations between e-cigarette use and e-cigarette flavors with cigarette smoking quit attempts and quit success: evidence from a U.S. large, nationally representative 2018-2019 survey. Nicotine Tob Res 2023 Feb 09;25(3):541-552 [FREE Full text] [doi: 10.1093/ntr/ntac241] [Medline: 36250607]
29. Selya A, Kim S, Shiffman S, Goldenson NI. Association of use of menthol- versus tobacco-flavored ENDS with switching completely away from cigarettes and differences by menthol cigarette smoking. Subst Use Misuse 2025;60(3):311-318. [doi: 10.1080/10826084.2024.2422963] [Medline: 39491336]
30. Shiffman S, Goldenson NI. Changes in dependence over one year among US adults who smoke cigarettes and switched completely or partially to use of the JUUL-brand electronic nicotine delivery system. Drug Alcohol Depend Rep 2023 Mar;6:100137 [FREE Full text] [doi: 10.1016/j.dadr.2023.100137] [Medline: 36994368]

XSL•FO
RenderX

31. Shiffman S, Goldenson NI, Hatcher C, Augustson EM. Changes in dependence as smokers switch from cigarettes to JUUL in two nicotine concentrations. Am J Health Behav 2021 May 01;45(3):563-575. [doi: 10.5993/AJHB.45.3.10] [Medline: 33894800]
32. Leavens EL, Nollen NL, Ahluwalia JS, Mayo MS, Rice M, Brett EI, et al. Changes in dependence, withdrawal, and craving among adult smokers who switch to nicotine salt pod-based e-cigarettes. Addiction 2022 Jan;117(1):207-215 [FREE Full text] [doi: 10.1111/add.15597] [Medline: 34105208]
33. Shiffman S, Oliveri DR, Goldenson NI, Liang Q, Black RA, Mishra S. Comparing adult smokers who switched to JUUL versus continuing smokers: biomarkers of exposure and of potential harm and respiratory symptoms. Nicotine Tob Res 2024 Mar 22;26(4):494-502. [doi: 10.1093/ntr/ntad197] [Medline: 37837438]
34. Cohen G, Goldenson NI, Bailey PC, Chan S, Shiffman S. Changes in biomarkers of cigarette smoke exposure after 6 days of switching exclusively or partially to use of the JUUL system with two nicotine concentrations: a randomized controlled confinement study in adult smokers. Nicotine Tob Res 2021 Nov 05;23(12):2153-2161 [FREE Full text] [doi: 10.1093/ntr/ntab134] [Medline: 34161586]
35. Holt NM, Shiffman S, Black RA, Goldenson NI, Sembower MA, Oldham MJ. Comparison of biomarkers of exposure among US adult smokers, users of electronic nicotine delivery systems, dual users and nonusers, 2018-2019. Sci Rep 2023 May 05;13(1):7297 [FREE Full text] [doi: 10.1038/s41598-023-34427-x] [Medline: 37147399]
36. Anic GM, Rostron BL, Hammad HT, van Bemmel DM, Del Valle-Pinero AY, Christensen CH, et al. Changes in biomarkers of tobacco exposure among cigarette smokers transitioning to ENDS use: the population assessment of tobacco and health study, 2013-2015. Int J Environ Res Public Health 2022 Jan 27;19(3):1462 [FREE Full text] [doi: 10.3390/ijerph19031462] [Medline: 35162490]
37. Chang JT, Anic GM, Rostron BL, Tanwar M, Chang CM. Cigarette smoking reduction and health risks: a systematic review and meta-analysis. Nicotine Tob Res 2021 Mar 19;23(4):635-642. [doi: 10.1093/ntr/ntaa156] [Medline: 32803250]

Abbreviations

AE: adverse event
ENDS: electronic nicotine delivery systems
OR: odds ratio

Edited by T de Azevedo Cardoso; submitted 03.07.24; peer-reviewed by T Smith, R Polosa; comments to author 09.12.24; revised version received 28.01.25; accepted 28.02.25; published 26.03.25.

Please cite as:
Goldenson NI, Shiffman S, Sembower MA, Selya A, Pype S, Black RA
Evaluating the Effect of the JUUL2 System With 5 Flavors on Cigarette Smoking and Tobacco Product Use Behaviors Among Adults Who Smoke Cigarettes: 6-Week Actual Use Study
Interact J Med Res 2025;14:e60620
URL: https://www.i-jmr.org/2025/1/e60620
doi:10.2196/60620
PMID:

©Nicholas I Goldenson, Saul Shiffman, Mark A Sembower, Arielle Selya, Steve Pype, Ryan A Black. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.03.2025. 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 Mediating Role of Depression in the Effect of Psychological Well-Being on the Self-Rated Health and Quality of Life of Older Adults: Cross-Sectional Study

Meng Zhang, MD; Ping Zhang, MD; Qiu Xiang, BS; Xiaopan Li*, MD; Suqin Xu*, BS

Nursing Department, Tongji Hospital affiliated to Tongji Medical College of Huazhong University of Science and Technology, No. 1095 Jie Fang Avenue, Hankou, Wuhan, P.R. China, China

*these authors contributed equally

Corresponding Author:
Suqin Xu, BS
Nursing Department, Tongji Hospital affiliated to Tongji Medical College of Huazhong University of Science and Technology, No. 1095 Jie Fang Avenue, Hankou, Wuhan, P.R. China, China

Abstract

Background: A growing body of evidence has identified that people's physical health could influence self-rated health and quality of life (QoL). However, only focusing on physical health is not adequate for the well-being of older adults. Studies focusing on the impact of psychological well-being on self-rated health and QoL are still rare.

Objective: This study aimed to identify the mediating effect of depression on the association between psychological well-being and self-rated health and QoL to comprehensively understand the relationship between them.

Methods: We used a cross-sectional study design and secondary data analysis from the Chinese Longitudinal Healthy Longevity Survey of 2017 to 2018. Path analysis was applied to examine the research questions.

Results: A large sample of 8839 older adults was included. Among them, more positive affect was found among those who were younger and had more years of schooling, higher household income, greater social security and social insurance, lower depression levels, and higher self-rated health levels. Depression had a partial mediation effect of psychological well-being on self-rated health and QoL, which explained 36% of the total variance (R^2=0.36). In addition, psychological well-being had a statistically significant direct effect on self-rated health and QoL (β=0.290; P<.001).

Conclusions: Our results indicate that psychological well-being had both direct and indirect effects on self-rated health and QoL. Depression was an important mediator that regulated the effect pathway in older adults.

(Interact J Med Res 2025;14:e57731) doi:10.2196/57731

KEYWORDS

aged; depression; health; quality of life; psychological well-being

Introduction

Populations are aging all over the world. The United Nations General Assembly announced in May 2020 that 2021 to 2030 would be the decade of healthy aging, emphasizing the need for worldwide policymakers to concentrate their efforts on enhancing the lives of older people now and in the future [1]. It is projected that the number of older adults worldwide will rise by approximately 20.6% by 2050, reaching an estimated 2 billion [2]. China has the largest population of older adults [3]. The percentage of Chinese citizens over 60 years of age is expected to reach 28% by 2040 [4]. The second baby boomers, those born between 1962 and 1975, will begin to retire in 2022, adding to China's aging problems [1].

Depression has become a prevalent mental health symptom among older adults [5,6]. Depression can be divided into clinical and nonclinical subtypes [7]. The clinical form, formally termed major depressive disorder, is a severe psychiatric syndrome marked by enduring dysphoria, anhedonia, and pervasive hopelessness [7]. In contrast, nonclinical depressive states exhibit milder symptoms and are usually precipitated by specific stressors or life events [7]. Community surveys indicate that the global prevalence of nonclinical depression in older adults ranges from 28.4% to 35.1%, whereas prevalence ranges from 27% to 37.3% in mainland China [5,6]. Epidemiological data further show that the frequency of clinical depression among older adults is 1.6% in the United States and 2.7% in China [8].

The course of depression in older adults tends to be more prolonged, and the therapeutic outcomes are generally less favorable compared to those of younger individuals [9]. Depression lowers quality of life (QoL) and raises the mortality risk of older adults [10]. In one study, compared to those without depression, older adults with depression reported lower QoL

XSL•FO
RenderX

[11]. A worse overall QoL was linked to a higher degree of depression regardless of the method used to measure QoL. The connections seemed to be steady over time [11].

Focusing on the psychological well-being of older people is essential to constructing a positive aging society and realizing the strategic goal of healthy aging. Healthy aging is defined as the process of developing and maintaining the functional ability that enables well-being in older age [12]. Self-rated questions are frequently used to evaluate health status and QoL, which are essential components of healthy aging [13,14].

Self-rated health (SRH) is the subjective assessment of an individual's current state of health. It is a valid and reliable measure among those without cognitive impairment [15]. Despite its seeming simplicity and subjectivity, SRH is an effective health indicator that predicts future health outcomes, including mortality, chronic disease incidence, and health care resource use [16]. SRH has also been shown to be stable across cultures, communities, and different age groups, and it has even been shown to be a more accurate predictor of all-cause death than tools created especially for this purpose [17]. A growing body of research reports that even after adjusting for demographic, social, and medical risk variables, SRH has been repeatedly demonstrated to be an independent predictor of mortality, exhibiting high reliability and validity as well as predictive value across a variety of health outcomes [17-19].

Self-rated QoL is one of the main elements that indicate healthy aging. QoL refers to the evaluation of the positive aspects of one's own life [20]. It is closely associated with social well-being and overall health; however, it is common for older adults to experience a decline in QoL as they age [20]. A recent study discussed a well-being paradox phenomenon that older adults with higher levels of SRH have worse health based on objective markers [21]. Another study pointed out that older adults may mitigate the disparity between potential and actual circumstances through the recalibration of expectations and internal benchmarks [22]. Therefore, they may experience improved QoL even if there are no appreciable health effects.

Prior studies have demonstrated that depression and anxiety mediate the relationship between social and biological factors and QoL [23]. Another longitudinal study indicated that increased depressive symptoms are associated with reduced SRH among Chinese individuals aged ≥45 years [24]. Most existing studies have primarily concentrated on social, biological, and psychopathological factors (depression and anxiety) as key determinants of SRH and QoL [23,24], thereby neglecting the underlying mechanisms that explain the connections between psychological variables (positive and negative affect) and SRH and QoL [25,26]. The mediating role of the main psychopathological factor (depression) in the relationship between psychological well-being and SRH and QoL remains unclear.

According to the hierarchy of needs theory by Maslow [27], humans will pursue higher-level needs only after their lower-level needs have been satisfied. We hypothesize that positive psychological well-being and the absence of emotional disorders are essential to the pursuit of self-actualization to obtain high SRH and QoL. The assumed relationships were that psychological well-being is positively associated with SRH and QoL, while depression is negatively associated with SRH and QoL. Based on this theory, we explored whether depression was a potential mediator of the association between psychological well-being and SRH and QoL in older people, and we applied further path analysis to clarify its potential mechanism. By examining this mediating effect, we can gain a comprehensive understanding of the complex interplay between these variables and potentially identify targets for intervention to improve the health and QoL of older adults. We obtained the data from a representative national survey and analyzed the statistics using path analysis. Our efforts may provide new insights into the mediating role of psychological well-being in promoting SRH and QoL in older people. We also hope to contribute to the body of knowledge on psychological well-being and its implications for the nursing care of older populations, ultimately aiming to promote positive aging.

Methods

Data and Study Design

The Chinese Longitudinal Healthy Longevity Survey (CLHLS) was launched in 1998 and continued through waves conducted in 2000, 2002, 2005, 2008-2009, 2011-2012, 2014, and 2017-2018 [28]. This cross-sectional study used data from the 2017-2018 survey wave of the CLHLS to undertake a secondary analysis. The CLHLS national survey was carried out in a randomly selected half of the counties and cities across 22 of the 31 provinces in China, which account for approximately 85% of the country's total population [28]. Nine provinces were excluded from the CLHLS, all located in western and northwestern China, where minority populations are comparatively large and age reporting tends to be less accurate [28]. The eligibility criteria for the CLHLS study were that the participants be aged 65 years or older [28]. The CLHLS tried to interview all centenarians who voluntarily agreed to participate in the study in the sampled counties and cities [28]. A targeted random sampling method was adopted to ensure the representativeness of the sample [28]. Specifically, it involved interviewing approximately equal numbers of men and women in their 90s, 80s, 70s, and 60s who lived in proximity to the centenarians [29]. We combined the data with complete statistics of demographics, a psychological well-being scale, the Center for Epidemiologic Studies Depression (CES-D) scale, SRH, and self-rated QoL question in the CLHLS survey. After excluding invalid questionnaires, the final sample size was 8839. The sample formula for path analysis was as follows: n=Z2P(1-P)d2, where *n* is the sample size, *Z* is a standard normal variate, *P* is the expected prevalence, and *d* represents precision or absolute error [30]. *P* was estimated based on the available literature, and its value ranged from 27% to 37.3% [5,6]. The admissible error (*d*) was set at 5% because the type I error rate of 0.05 was acceptable for the difference between the sample statistics and the population parameter to be no more than 5% of the population parameter with a standard 95% confidence level (*Z*=1.96) [30]. The values used in this calculation were *Z*=1.96, *P*=.27 to .37, and *d*=0.05. Accounting for a 20% loss to follow-up rate, the estimated sample size ranged from 362

XSL•FO
RenderX

to 429. Finally, the minimum sample size in this study was set at 429.

Variables

Independent Variable

Psychological well-being was measured using 7 items from the 1998 CLHLS, including positive affect and negative affect [31,32]. Positive affect was assessed using 4 positive items (optimistic: "Do you always look on the bright side of things?"; conscientious: "Do you like to keep your belongings neat and clean?"; sense of personal control: "Can you make your own decisions concerning your personal affairs?"; a positive attitude toward one's aging: "Are you as happy as when you were younger?"), and negative affect was assessed using 3 negative items (agitation: "Do you often feel fearful or anxious?"; loneliness: "Do you often feel lonely and isolated?"; perceived loss of self-worth: "Do you feel the older you get, the more useless you are?"), which were rated on a 5-point frequency scale from 0 ("never") to 4 ("always") in both cases [28]. After reverse coding the negative items, the sum of the scores for all 7 items ranged from 0 to 28, with a higher score indicating more positive psychological well-being [28]. The Cronbach α reached 0.704 to 0.752 in previous studies, which indicates an acceptable internal consistency [31-34].

Mediation Variable

The 10-item CES-D scale was developed by Andresen et al [35] in 1994 for use with older adults who were relatively well. The CES-D is a screening tool to identify individuals and population groups at elevated risk of nonclinical depression [36]. The internal consistency was satisfactory (Cronbach α=0.78-0.79), and moderate consistency over 3 years was also found to be significant (r=0.44; P<.01) [37]. The CLHLS used Chinese translations available at the Center for Epidemiologic Studies website [28]. Interviewees had 5 answers to choose from in the 10 CES-D items: "never," "rarely," "sometimes," "often," and "always." For questions 1, 2, 3, 4, 6, 8, and 9, we scored 3 for "always" and 0 for "never." For questions 5, 7, and 10, we scored 0 for "always (good)" and 3 for "never (very bad)" [28]. Therefore, CES-D scores ranged from 0 to 30, with 10 and 20 as cutoff points for the levels of depressive symptoms [38].

Dependent Variables

SRH, introduced in the 1980s, is a subjective indicator of health status for older adults [39]. It was measured using the following question: "How do you rate your life at present?" [39]. Self-rated QoL is a subjective assessment of QoL by individuals [40]. It was evaluated using the following question: "How would you rate your overall quality of life?" [40]. SRH and QoL were rated on a 4-point scale ranging from 0 ("very bad") to 4 ("very good") [39,40]. Higher scores represented better SRH and QoL status. The widely used single-question method has been found to be reliable and valid [13,41,42].

Data Analysis

SPSS Statistics (version 25.0; IBM Corp) and SPSS Amos (version 23.0; IBM Corp) were used to analyze the data. Continuous variables were represented using means and SDs or medians and IQRs, whereas categorical or ranked variables were represented using frequencies and percentages. The chi-square test was conducted to assess categorical data, and the Kruskal-Wallis test was conducted to evaluate continuously ranked data. The Spearman correlation was used to examine the relationship among sociodemographic variables, depression, SRH, and QoL. The hypothesized model was tested using the maximum likelihood method in the SPSS Amos statistical software. Several model fit indices were estimated to test the good fit of the model, including the chi-square test, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), cumulative fit index (CFI), incremental fit index, goodness-of-fit index (GFI), normed fit index (NFI), and Tucker-Lewis index (TLI). A good fit model was considered when all path coefficients were significant at the level of .05, χ^2<5, SRMR<0.08, RMSEA<0.08, GFI>0.9, NFI>0.9, TLI>0.9, and CFI>0.9 [43]. All analyses were 2-tailed.

Ethical Considerations

The CLHLS study was approved by the Duke Health Institutional Review Board (Pro00062871) and the Research Ethics Committee of Peking University (IRB00001052-13074), and all participants provided written informed consent. The data were made publicly available for secondary analysis after being fully anonymized. No compensation was provided to participants.

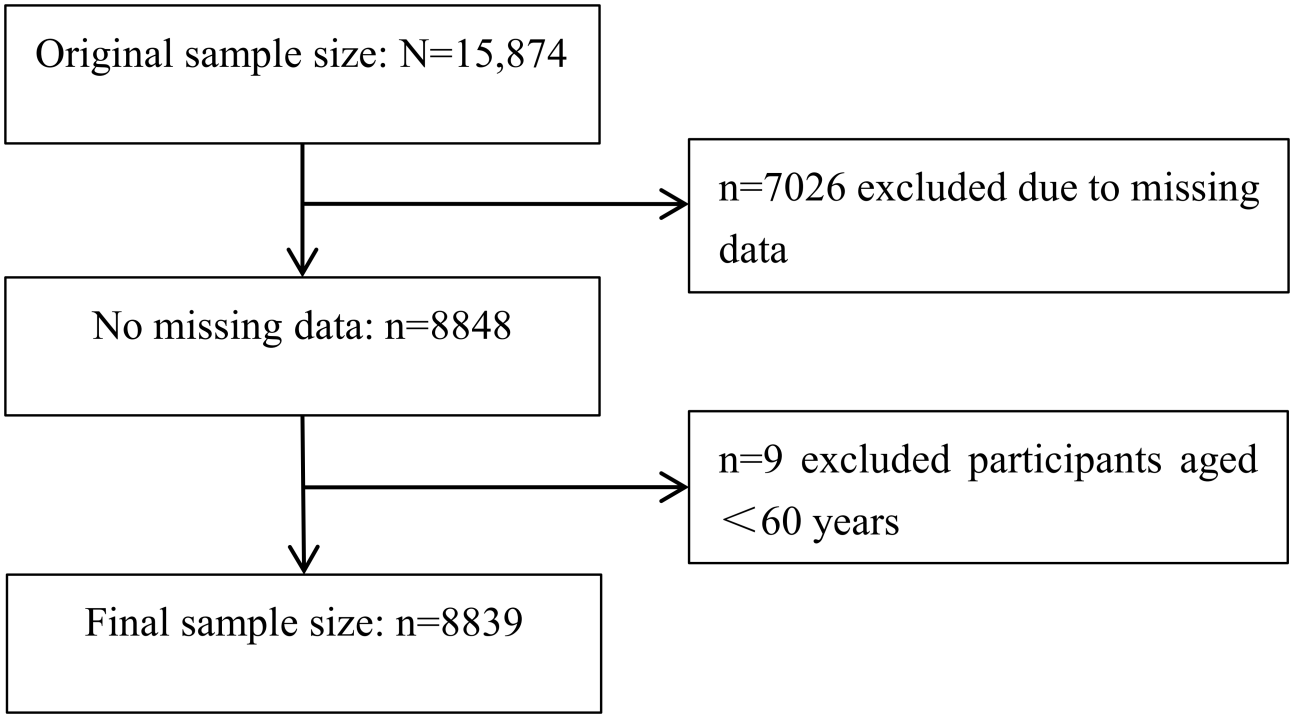
Results

Participant Demographics

A total of 8839 older adults were included in this study (Figure 1), with 48.7% (4305/8839) being male and 51.3% (4534/8839) being female. The average age was 81.27 (SD 10.29) years, and they had a median of 3 (IQR 0-6) years of schooling. The median total income in the previous year was ¥30,000 (US $4212.57; IQR ¥8000 to >¥99998 [US $1123.35 to >US $14,041.62]). Approximately half (4384/8839, 49.6%) of the participants were currently married or living with a partner. More than 90% of the participants (8140/8839, 92.1%) had social security or social insurance. The average scores for depression, SRH, and QoL were 9.62, 2.51, and 2.91, respectively. A total of 52.1% (4607/8839) of the participants scored lower than 10 in the depression scale, 45.5% (4028/8839) scored between 10 and 20, and 2.3% (204/8839) scored higher than 20.

XSL•FO
RenderX

Figure 1. Flowchart of the sample selection.



Variables According to Different Psychological Well-Being Levels

A substantial difference was observed in age, sex, years of schooling, total household income in the previous year, social security and social insurance status, depression, QoL, and SRH (Table 1). Younger participants and those who had more years of schooling, higher household income, social security and social insurance, lower depression levels, and higher SRH and QoL levels indicated more positive affect. Female participants had higher positive affect levels than male participants.

Table . Variables according to different psychological well-being levels. Psychological well-being was divided into 3 levels based on tertile range.

Variable	Psychological well-being level			Statistic	
	More negative (score of <19; n=3832)	Moderate (score of 19-21; n=2124)	More positive (score of >21; n=2883)	Kruskal-Wallis *H* statistic	Chi-square (*df*)
Age (y), mean (SD)	81.72 (10.32)	80.77 (10.29)	81.02 (10.23)	13.72[a]	—[b]
Sex (female), n (%)	2063 (53.84)	1023 (48.16)	1448 (50.23)	—	19.6[c] (2)
Years of schooling, median (IQR)	2 (0 to 6)	3 (0 to 6)	4 (0 to 8)	76.55[c]	—
Total income in the previous year, median (IQR)	30,000 (US $4212.57; IQR ¥6000 to 80,000 [US $842.51 to US $11,233.52])	35,000 (US $4914.67; IQR ¥9000 to 99998 [US $1263.77 to US $14,041.62])	40,000 (US $5616.76; IQR ¥10,000 to 99998 [US $1404.19 to US $14,041.62])	54.37[c]	—
Social security and insurance (yes), n (%)	3490 (91.1)	1959 (92.23)	2691 (93.34)	—	9.5[a] (2)
Depression, mean (SD)	12.36 (1 to 29)	9.16 (1 to 28)	6.32 (1 to 26)	2578.38[c]	—
QoL[d], mean (SD)	2.68 (1 to 4)	2.92 (1 to 4)	3.21 (1 to 4)	773.07[c]	—
Self-rated health, mean (SD)	2.23 (1 to 4)	2.53 (1 to 4)	2.86 (1 to 4)	955.07[c]	—

[a]*P*<.01.

[b]Not applicable.

[c]*P*<.001.

[d]QoL: quality of life.

XSL•FO
RenderX

Correlation of Variables

Table 2 shows the correlation among psychological well-being, depression, SRH, and QoL. The Spearman correlation suggests that psychological well-being had a negative correlation with depression (β=–0.590) but psychological well-being had a positive correlation with QoL (β=0.326) and SRH (β=0.361). Moreover, depression was significantly negatively correlated with QoL (β=–0.354) and SRH (β=–0.382). In addition, QoL was significantly positively correlated with SRH (β=0.463).

Table . Spearman correlation of variables.

Variable	Psychological well-being (*r*)	Depression (*r*)	QoL[a] (*r*)	Self-rated health (*r*)
Psychological well-being (*r*)	1	—[b]	—	—
Depression (*r*)	−0.590[c]	1	—	—
QoL (*r*)	0.326[c]	−0.354[c]	1	—
Self-rated health (*r*)	0.361[c]	−0.382[c]	0.463[c]	1

[a]QoL: quality of life.

[b]Not applicable.

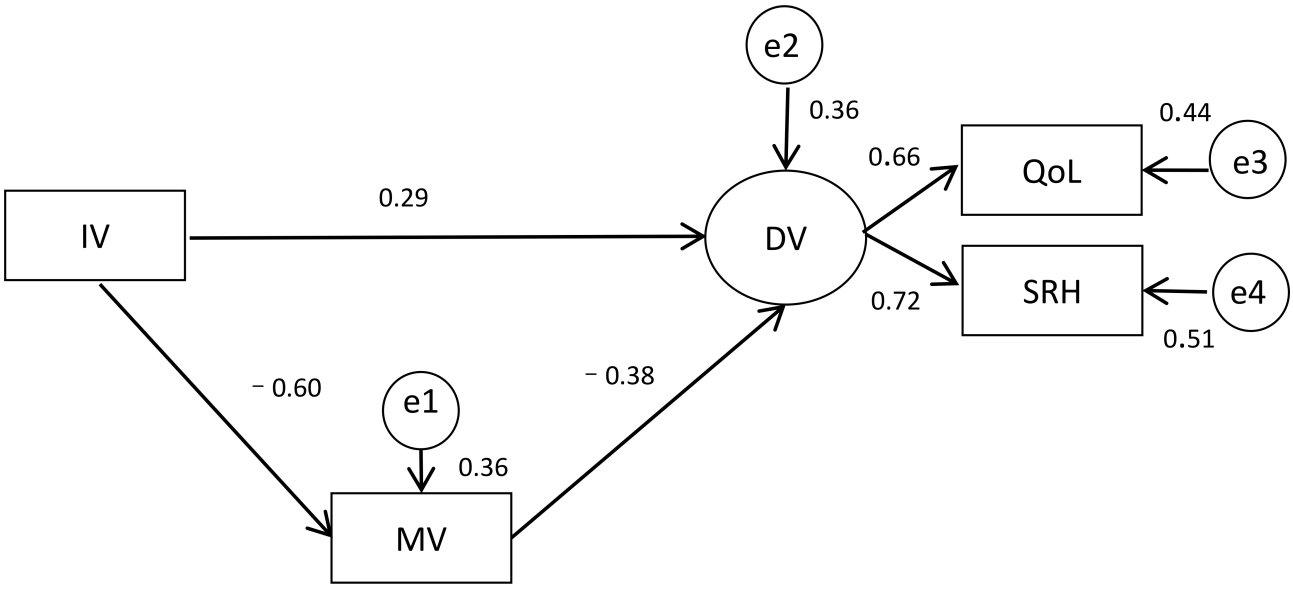
[c]*P*<.001.

Mediator Model

Figure 2 illustrates the standardized effects of the coefficients in the mediator model. All path estimates were statistically significant (*P*<.001) and correlated with the hypothesized model. Depression had a partial mediation effect of psychological well-being on SRH and QoL, which explained 36% of the total variance (R^2=0.36). Psychological well-being had a statistically significant negative direct effect on depression (β=−0.603; *P*<.001), whereas depression had a further statistically significant negative direct effect on SRH and QoL (β=–0.378; *P*<.001). In addition, psychological well-being had a statistically significant positive direct effect on SRH and QoL (β=0.290; *P*<.001). Furthermore, the indirect effect of psychological well-being on general health was β = (−0.603) × (−0.378) = 0.228. In addition, SRH (β=0.716) contributed more to general health than QoL (β=0.663).

Fit indexes of the mediator model revealed a good fit of the data (χ^2_1=1.7; *P*=.19). The SRMR was 0.006, and the RMSEA was 0.009, which indicated good validity of the measurement construct. The scores of the GFI, NFI, TLI, CFI, and incremental fit index were over 0.9, further illustrating the good model fit.

Figure 2. Mediator model. DV: dependent variable (quality of life [QoL] and self-rated health [SRH]); e: the residual of the variable; IV: independent variable (psychological well-being); MV: mediation variable (depression).



Discussion

Principal Findings

Our study found that psychological well-being had a negative direct effect on depression, whereas depression had a further negative direct effect on SRH and QoL. In addition, psychological well-being had a positive direct effect on SRH and QoL. The results were consistent with those of previous research that highlighted the effect of psychological well-being and depression on SRH and QoL [33]. Moreover, the results also illustrated that younger older adults who had more years of schooling, higher household income, and greater social security and social insurance indicated more positive affect. These findings expand those of previous studies by exploring

XSL•FO
RenderX

the mediating role of depression among older adults in China, thus providing a reference for improving the psychological well-being of Chinese older adults.

Psychological well-being was regulated by older people's demographics, such as age, sex, educational background, and income. A very interesting finding was that people aged >81 years tended to have more positive affect or more negative affect. This could be explained by the "response shift" phenomenon, in which some older adults reconcile with the deterioration of health and functionality [22]. This reconciliation often involves the recalibration of expectations and internal benchmarks to mitigate the disparity between potential and actual circumstances through both biological and psychosocial transformations [22]. Our study found that women had more negative affect than men. The reason may be that women have more negative reactions when encountering negative stimuli in comparison to men [44]. The results also demonstrated that total income, years of schooling, and social security and insurance status were positively associated with psychological well-being. Previous studies have proved the modest correlations between socioeconomic status, such as income, educational level, and occupation, and personal psychological well-being, which is consistent with the results of our study [45].

This study indicated that positive affect decreased the incidence of depression and enhanced the SRH and QoL of older adults. Our results were aligned with those of a previous study [33] and provided a specific pathway and coefficient statistics among variables of psychological well-being, depression, SRH, and QoL. Psychological well-being is very important to an individual and may predict the occurrence of late-life depression in older adults. The older adults who had more positive affect were relatively more likely to look at the bright side of the world and focus on the goodness of things [46]. In positive psychology theory, the novel ideas and actions (eg, the urge to play and explore) brought by broadened mindsets have been proven to alleviate depression and improve general health [47].

The path analysis revealed that depression was a mediator of the effect of psychological well-being on SRH and QoL in older adults, which accounted for 36% of the variance. The size of its indirect effect of psychological well-being on SRH and QoL (β=0.228) was very close to the direct effect size (β=0.290), which implied that depression was an important mediator between psychological well-being and SRH and QoL. Depression is a mental health symptom and has adverse consequences for SRH and QoL in older adults. Approximately half (4232/8839, 47.9%) of the older adults in this study had CES-D scores of 10 and over, higher than the incidence in Indonesia and previous samples in China, such as the 2011 to 2012 wave [48]. Our findings are consistent with the hypothesis that depressive symptoms among Chinese older adults have risen in recent years [49]. Nevertheless, focusing on the psychological well-being of older adults is urgent to promote healthy aging.

Paying attention to psychological well-being and reducing depression may be an effective strategy for promoting the health and QoL of older people. Previous studies have indicated that people with depression show specific deficits in being able to anticipate future positive events [50]. The ability to think mindfully, develop a tolerant outlook, please oneself, or cope when confronting situations that lead to negative affect may be powerful measures to construct healthy psychological well-being. We should also focus on the support systems of older adults, which serve as a protective factor for psychological well-being to promote the health of older adults [51].

China has the fifth highest older adult population worldwide, which is frequently cited as an example for other middle-income nations [52]. Our results could also have implications for limited-income countries. Future research could involve additional countries to compare the influence of various cultures and health care systems on older adults' psychological well-being, health, and QoL.

Limitations

Although our data were derived from a large Chinese population-based older adult cohort, there were still some limitations to our study. First, the cross-sectional design could not explain the causal effects of the variables when compared with longitudinal studies. These results may not be temporally stable over time. Second, the participants may have provided biased responses due to memory recall issues when the data were collected. Third, our study only examined the relationships between 4 indicators: psychological well-being, depression, SRH, and QoL. If more indicators had been included, the conclusions may have been richer. Fourth, the analysis was based on self-reported questions on depression, well-being, and QoL. Moreover, the CES-D is a screening tool, not a clinical diagnostic tool.

Conclusions

Our study provides a new insight into the influence of the psychological aspect on older adults' health and QoL. We also discussed the variable of depression, a very common type of emotional distress, in the mediation pathway between psychological well-being and SRH and QoL. Depression was an important mediator of the effect of psychological well-being on SRH and QoL in older people. More attention should be paid to depression and the psychological well-being of older adults. Future tailored interventions should focus on relieving depression in older adults and promoting their psychological well-being.

Conflicts of Interest

None declared.

References

XSL•FO
RenderX

1. Chen X, Giles J, Yao Y, et al. The path to healthy ageing in China: a Peking University–Lancet Commission. Lancet 2022 Dec 3;400(10367):1967-2006. [doi: 10.1016/S0140-6736(22)01546-X] [Medline: 36423650]
2. World population prospects 2019: highlights. United Nations Department of Economic and Social Affairs. 2019. URL: https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf [accessed 2025-10-22]
3. Bai C, Lei X. New trends in population aging and challenges for China's sustainable development. China Econ J 2019;13(1):3-23. [doi: 10.1080/17538963.2019.1700608]
4. Ageing and health in China. World Health Organization. URL: https://www.who.int/china/health-topics/ageing [accessed 2025-10-22]
5. Wang Y, Liu M, Yang F, Chen H, Wang Y, Liu J. The associations of socioeconomic status, social activities, and loneliness with depressive symptoms in adults aged 50 years and older across 24 countries: findings from five prospective cohort studies. Lancet Healthy Longev 2024 Sep;5(9):100618. [doi: 10.1016/j.lanhl.2024.07.001] [Medline: 39208829]
6. Zhu L, Wang Y, Li J, Zhou H, Li N, Wang Y. Depressive symptoms and all-cause mortality among middle-aged and older people in China and associations with chronic diseases. Front Public Health 2024 May 22;12:1381273. [doi: 10.3389/fpubh.2024.1381273] [Medline: 38841667]
7. Parker G, Paterson A. Differentiating "clinical" and "non-clinical" depression. Acta Psychiatr Scand 2015 Jun;131(6):401-407. [doi: 10.1111/acps.12385] [Medline: 25565360]
8. Wang F, Zhang QE, Zhang L, et al. Prevalence of major depressive disorder in older adults in China: a systematic review and meta-analysis. J Affect Disord 2018 Dec 1;241:297-304. [doi: 10.1016/j.jad.2018.07.061] [Medline: 30142588]
9. Haigh EA, Bogucki OE, Sigmon ST, Blazer DG. Depression among older adults: a 20-year update on five common myths and misconceptions. Am J Geriatr Psychiatry 2018 Jan;26(1):107-122. [doi: 10.1016/j.jagp.2017.06.011] [Medline: 28735658]
10. Shrira A. The effect of lifetime cumulative adversity on change and chronicity in depressive symptoms and quality of life in older adults. Int Psychogeriatr 2012 Dec;24(12):1988-1997. [doi: 10.1017/S1041610212001123] [Medline: 22874666]
11. Sivertsen H, Bjørkløf GH, Engedal K, Selbæk G, Helvik AS. Depression and quality of life in older persons: a review. Dement Geriatr Cogn Disord 2015;40(5-6):311-339. [doi: 10.1159/000437299] [Medline: 26360014]
12. Beard JR, Officer A, de Carvalho IA, et al. The world report on ageing and health: a policy framework for healthy ageing. Lancet 2016 May 21;387(10033):2145-2154. [doi: 10.1016/S0140-6736(15)00516-4] [Medline: 26520231]
13. Zou M, Sun C, Yang M, et al. Influences of tea consumption on self-rated health and life satisfaction among older adults: evidence from the CLHLS. Psych J 2025 Feb;14(1):62-73. [doi: 10.1002/pchj.807] [Medline: 39462173]
14. Fu X, Du B, Chen Q, et al. Self-rated health (SRH), recovery from work, fatigue, and insomnia among commercial pilots concerning occupational and non-occupational factors. Front Public Health 2023 Jan 4;10:1050776. [doi: 10.3389/fpubh.2022.1050776] [Medline: 36684912]
15. Bombak AE. Self-rated health and public health: a critical perspective. Front Public Health 2013 May 20;1:15. [doi: 10.3389/fpubh.2013.00015] [Medline: 24350184]
16. Chan YY, Teh CH, Lim KK, et al. Lifestyle, chronic diseases and self-rated health among Malaysian adults: results from the 2011 National Health and Morbidity Survey (NHMS). BMC Public Health 2015 Aug 6;15:754. [doi: 10.1186/s12889-015-2080-z] [Medline: 26246019]
17. Ganna A, Ingelsson E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. Lancet 2015 Aug 8;386(9993):533-540. [doi: 10.1016/S0140-6736(15)60175-1] [Medline: 26049253]
18. Breidablik HJ, Meland E, Holmen TL, Lydersen S. Role of parents in adolescent self-rated health: Norwegian Nord-Trøndelag health study. Adolesc Health Med Ther 2010 Sep 6;1:97-104. [doi: 10.2147/AHMT.S12877] [Medline: 24600265]
19. Jylhä M. What is self-rated health and why does it predict mortality? Towards a unified conceptual model. Soc Sci Med 2009 Aug;69(3):307-316. [doi: 10.1016/j.socscimed.2009.05.013] [Medline: 19520474]
20. Huang X, Zhang M, Fang J, Zeng Q, Wang J, Li J. Classifying and characterizing the development of self-reported overall quality of life among the Chinese elderly: a twelve-year longitudinal study. BMC Public Health 2022 Jun 7;22(1):1139. [doi: 10.1186/s12889-022-13314-6] [Medline: 35672842]
21. Whitmore C, Markle-Reid M, McAiney C, et al. Self-reported health and the well-being paradox among community-dwelling older adults: a cross-sectional study using baseline data from the Canadian Longitudinal Study on Aging (CLSA). BMC Geriatr 2022 Feb 10;22(1):112. [doi: 10.1186/s12877-022-02807-z] [Medline: 35144559]
22. Andreas S, Schulz H, Volkert J, et al. Prevalence of mental disorders in elderly people: the European MentDis_ICF65+ study. Br J Psychiatry 2017 Feb;210(2):125-131. [doi: 10.1192/bjp.bp.115.180463] [Medline: 27609811]
23. Susanti H, Fajri M, Keliat BA, Brooks H, Bee P, Maharani A. Exploring the mediating role of depression and anxiety in the relationship between social and biological factors and quality of life in Indonesia: a structural equation modelling approach. BMJ Open 2025 Apr 28;15(4):e095110. [doi: 10.1136/bmjopen-2024-095110] [Medline: 40295135]
24. Leng S, Yao L, Deng J. Associations between self-rated health and depressive symptoms among middle-aged and older adults in China: a cross-lagged panel analysis (2011-2020). PLoS ONE 2025 Apr 29;20(4):e0321272. [doi: 10.1371/journal.pone.0321272] [Medline: 40299937]
25. Lopez J, Perez-Rojo G, Noriega C, Sánchez-Cabaco A, Sitges E, Bonete B. Quality-of-life in older adults: its association with emotional distress and psychological wellbeing. BMC Geriatr 2024 Oct 9;24(1):815. [doi: 10.1186/s12877-024-05401-7] [Medline: 39385087]

XSL•FO
RenderX

26. Lee MC, Tai CY, Liu CC, Wang WH, Chang CF, Chen MQ. The mediating role of psychological well-being in the relationship between self-management and quality of life in patients with hypertension. West J Nurs Res 2024 Feb;46(2):125-132. [doi: 10.1177/01939459231217934] [Medline: 38112109]
27. Hale AJ, Ricotta DN, Freed J, Smith CC, Huang GC. Adapting Maslow's Hierarchy of Needs as a Framework for Resident Wellness. Teach Learn Med 2019;31(1):109-118. [doi: 10.1080/10401334.2018.1456928] [Medline: 29708437]
28. Zeng Y, Feng Q, Hesketh T, Christensen K, Vaupel JW. Survival, disabilities in activities of daily living, and physical and cognitive functioning among the oldest-old in China: a cohort study. Lancet 2017 Apr 22;389(10079):1619-1629. [doi: 10.1016/S0140-6736(17)30548-2] [Medline: 28285816]
29. Zeng Y, Zheng Z. Longitudinal survey and interdisciplinary research on healthy aging. China Popul Dev Stud 2020 Apr;3:293-296. [doi: 10.1007/s42379-020-00051-9]
30. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med 2013 Apr;35(2):121-126. [doi: 10.4103/0253-7176.116232] [Medline: 24049221]
31. Zhou Z, Cai L, Zhuang M, Hong YA, Fang Y. A longitudinal analysis of the association between the living arrangements and psychological well-being of older Chinese adults: the role of income sources. BMC Geriatr 2019 Dec 10;19(1):347. [doi: 10.1186/s12877-019-1371-0] [Medline: 31822282]
32. Han WJ, Shibusawa T. Trajectory of physical health, cognitive status, and psychological well-being among Chinese elderly. Arch Gerontol Geriatr 2015;60(1):168-177. [doi: 10.1016/j.archger.2014.09.001] [Medline: 25277446]
33. Zhang L, Wei X, Ma X, Ren Z. Physical health and quality of life among older people in the context of Chinese culture. Int J Environ Res Public Health 2021 Jun 24;18(13):6798. [doi: 10.3390/ijerph18136798] [Medline: 34202846]
34. Yang W, Tian G, Cui Y, et al. Associations between activities of daily living, leisure activities and subjective well-being in Chinese older adults. Gerodontology 2024 Sep;41(3):415-423. [doi: 10.1111/ger.12721] [Medline: 37948355]
35. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). Am J Prev Med 1994;10(2):77-84. [Medline: 8037935]
36. Yu X, Jones RN, Kobayashi LC, Gross AL. Cross-national statistical harmonization of the Center for Epidemiologic Studies Depression (CES-D) scale among older adults in China, England, India, Mexico, South Africa, and the United States. J Clin Epidemiol 2025 Feb;178:111623. [doi: 10.1016/j.jclinepi.2024.111623] [Medline: 39642944]
37. Boey KW. Cross-validation of a short form of the CES-D in Chinese elderly. Int J Geriatr Psychiatry 1999 Aug;14(8):608-617 [FREE Full text] [doi: 10.1002/(sici)1099-1166(199908)14:8<608::aid-gps991>3.0.co;2-z] [Medline: 10489651]
38. Smith J, Gerstorf D, Li Q. Psychological resources for well-being among octogenarians, nonagenarians, and centenarians: differential effects of age and selective mortality. In: Yi Z, Poston DL, Vlosky DA, Gu D, editors. Healthy Longevity in China: Springer:329-345 URL: https://link.springer.com/chapter/10.1007/978-1-4020-6752-5_20 [doi: 10.1007/978-1-4020-6752-5_20]
39. Macias C, Gold PB, Öngür D, Cohen BM, Panch T. Are single-item global ratings useful for assessing health status? J Clin Psychol Med Settings 2015 Oct 22. [doi: 10.1007/s10880-015-9436-5] [Medline: 26492891]
40. Cheung F, Lucas RE. Assessing the validity of single-item life satisfaction measures: results from three large samples. Qual Life Res 2014 Dec;23(10):2809-2818. [doi: 10.1007/s11136-014-0726-4] [Medline: 24890827]
41. Xu Q, Jia C. A study on the effects of informal care on the health of older adults-evidence from China. Public Health Nurs 2024;41(5):1154-1164. [doi: 10.1111/phn.13386] [Medline: 39051737]
42. Fu L, Liu C, Dong Y, et al. Mediating effects of information access on internet use and multidimensional health among middle-aged and older adults: nationwide cross-sectional study. J Med Internet Res 2024 Sep 9;26:e49688. [doi: 10.2196/49688] [Medline: 39250790]
43. Byrne BM. Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, 3rd edition: Taylor & Francis; 2016.
44. Poláčková Šolcová I, Lačev A. Differences in male and female subjective experience and physiological reactions to emotional stimuli. Int J Psychophysiol 2017 Jul;117:75-82. [doi: 10.1016/j.ijpsycho.2017.04.009] [Medline: 28454989]
45. Navarro-Carrillo G, Alonso-Ferres M, Moya M, Valor-Segura I. Socioeconomic status and psychological well-being: revisiting the role of subjective socioeconomic status. Front Psychol 2020 Jun 10;11:1303. [doi: 10.3389/fpsyg.2020.01303] [Medline: 32587560]
46. Xu Y, Yu Y, Xie Y, et al. Positive affect promotes well-being and alleviates depression: the mediating effect of attentional bias. Psychiatry Res 2015 Aug 30;228(3):482-487. [doi: 10.1016/j.psychres.2015.06.011] [Medline: 26142834]
47. Donaldson SI, Csikszentmihalyi M, Nakamura J. Applied Positive Psychology: Improving Everyday Life, Health, Schools, Work and Society: Taylor & Francis; 2011.
48. Lei X, Sun X, Strauss J, Zhang P, Zhao Y. Depressive symptoms and SES among the mid-aged and elderly in China: evidence from the China Health and Retirement Longitudinal Study national baseline. Soc Sci Med 2014 Nov;120:224-232. [doi: 10.1016/j.socscimed.2014.09.028] [Medline: 25261616]
49. Hu X, Liu H, Liu Q, et al. Depressive symptoms and their influencing factors among older adults in China: a cross-sectional study. Front Public Health 2024 Nov 15;12:1423391. [doi: 10.3389/fpubh.2024.1423391] [Medline: 39618958]

XSL•FO
RenderX

50. Edmondson OJ, MacLeod AK. Psychological well-being and anticipated positive personal events: their relationship to depression. Clin Psychol Psychother 2015;22(5):418-425. [doi: 10.1002/cpp.1911] [Medline: 25044547]
51. Zhang C, Zhao Y, Li X, et al. The effect of self-rated health on depressive symptoms in Chinese older adults: the mediating role of social participation and spouse health. Geriatr Nurs 2024;59:411-417. [doi: 10.1016/j.gerinurse.2024.07.033] [Medline: 39141948]
52. Chen X, Giles J, Yao Y, et al. The path to healthy ageing in China: a Peking University-Lancet Commission. Lancet 2022 Dec 3;400(10367):1967-2006. [doi: 10.1016/S0140-6736(22)01546-X] [Medline: 36423650]

Abbreviations

CES-D: Center for Epidemiologic Studies Depression
CFI: cumulative fit index
CLHLS: Chinese Longitudinal Healthy Longevity Survey
GFI: goodness-of-fit index
NFI: normed fit index
QoL: quality of life
RMSEA: root mean square error of approximation
SRH: self-rated health
SRMR: standardized root mean square residual
TLI: Tucker-Lewis index

Edited by N Cahill; submitted 25.02.24; peer-reviewed by C Chen, N Ayadi, S Dargahi; revised version received 14.10.25; accepted 14.10.25; published 04.11.25.

Please cite as:
Zhang M, Zhang P, Xiang Q, Li X, Xu S
The Mediating Role of Depression in the Effect of Psychological Well-Being on the Self-Rated Health and Quality of Life of Older Adults: Cross-Sectional Study
Interact J Med Res 2025;14:e57731
URL: https://www.i-jmr.org/2025/1/e57731
doi:10.2196/57731

© Meng Zhang, Ping Zhang, Qiu Xiang, Xiaopan Li, Suqin Xu. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 4.11.2025. 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 Prevention Focus and Sedentary Behavior in Older Adults: Cross-Sectional Study

Jethro Raphael Suarez[1,2], MS; Amber Blount[3], BS; Kworweinski Lafontant[2,4], MS; Joon-Hyuk Park[1,5,6], PhD; Rui Xie[2,7], PhD; Nichole Lighthall[8], PhD; Ladda Thiamwong[2,6], RN, PhD

[1]Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL, United States

[2]College of Nursing, University of Central Florida, Orlando, FL, United States

[3]Department of Counselor Education and School Psychology, University of Central Florida, Orlando, FL, United States

[4]Institute of Exercise Physiology and Rehabilitation Science, University of Central Florida, Orlando, FL, United States

[5]Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology, 333 Techno Jungang-daero, Hyeonpung-eup, Dalseong-gun, Daegu, Republic of Korea

[6]Disability, Aging, and Technology Cluster, University of Central Florida, Orlando, FL, United States

[7]Department of Statistics and Data Science, University of Central Florida, Orlando, FL, United States

[8]Department of Psychology, University of Central Florida, Orlando, FL, United States

Corresponding Author:
Joon-Hyuk Park, PhD
Department of Mechanical and Aerospace Engineering, University of Central Florida, Orlando, FL, United States

Abstract

Background: Older adults engage in increased amounts of sedentary behavior (SB), which can result in a significant decline in muscle function and overall health. An understanding of the motivational driving factors that lead older adults to engage in SB can help to create effective intervention programs.

Objective: This study aimed to determine the association between prevention and promotion focus with SB in older adults, as well as compare these associations with two factors (ie, age and BMI) that are commonly known to have an association with SB among older adults.

Methods: A cross-sectional analysis was conducted among 93 community-dwelling older adults with a mean age of 74.98 (SD 6.68) years. Prevention and promotion focus were both assessed using the Regulatory Focus Questionnaire. Correlation analysis was performed to determine the associations between prevention focus, promotion focus, age, and BMI with SB. Anderson-Darling tests confirmed nonnormal data distributions for all factors (except age); therefore, Spearman rank correlation was used to determine correlations between factors. Comparative analysis of significant correlations was performed using Fisher Z transformation.

Results: Prevention focus had the greatest statistically significant correlation with SB (ρ=0.296; *P*=.004), followed by BMI (ρ=0.204; *P*=.049). Both age (ρ=0.116; *P*=.27) and promotion focus (ρ=0.002; *P*=.99) had statistically insignificant correlations with SB, indicating no associations. The correlation between prevention focus and SB did not significantly differ from the correlation between BMI and SB (*P*=.51).

Conclusions: Prevention focus was found to have a weak, but significant positive association with SB in older adults. Although age and BMI have been found to have an association with SB in previous literature, age was not associated with SB in this study, while BMI had a significant but relatively weaker association with SB than that with prevention focus. However, the association found between BMI and SB did not statistically differ from the association found between prevention focus and SB. These findings suggest that older adults could be driven to engage in increased amounts of SB due to having a dominant prevention focus, which revolves around thoughts of safety and avoiding negative consequences. The recognition of this association has the potential to aid in developing intervention programs that could promote shifting from prevention to promotion focus, thereby reducing SB in older adults.

Trial Registration: ClinicalTrials.gov NCT05778604; https://www.clinicaltrials.gov/study/NCT05778604

International Registered Report Identifier (IRRID): RR2-10.2196/51899

(Interact J Med Res 2025;14:e63280) doi:10.2196/63280

KEYWORDS

accelerometry; regulatory focus theory; actigraph; motivation; physical activity

XSL•FO
RenderX

Introduction

Sedentary behavior (SB) is most prevalent among older adults and negatively affects their physical health by increasing their risk for diseases, as well as overall mortality [1]. SB can be defined as any waking behavior that results in an energy expenditure of less than 1.5 metabolic equivalents while in a sitting or reclining position [2-4]. Various intervention programs that aim to reduce SB in older adults have been developed and focus primarily on changing psychological and behavioral factors, rather than environmental or physical factors [5]. These interventions include self-regulatory strategies, which include goal setting and behavioral feedback with participants [6-8]. Self-regulation is described as the process of aligning an individual's behavior and decision-making with their goals [9]. Regulatory Focus Theory builds on the idea of self-regulation by suggesting that individuals have two distinct motivational orientations that influence their decision-making and behavior: promotion and prevention focus [9-11]. Individuals with a prevention focus strive for feelings of safety and responsibility, and seek to minimize the negative outcomes or consequences in their lives (i.e., avoid losses) [10,11]. Conversely, individuals with a promotion focus strive for growth and advancement, seek opportunities for success, aim to maximize potential gains, and strive to obtain positive outcomes in life [10,11].

Regulatory focus theory has been found to relate to older adults, as evidenced by a positive relationship between negative self-perceptions on aging (ie, a pessimistic or unfavorable view toward aging) and trait prevention focus [12]. Negative self-perceptions of aging in older adults have also been found to have a negative relationship with self-reported physical activity (PA) and performance in daily activities [13-15]. This is likely due to a belief that PA and overall movement are not beneficial in the long term and could increase their risk of falling. Although regulatory focus theory has not yet been directly linked to SB in older adults, it has the potential to be related to SB, due to the previously found relationships with negative self-perceptions of aging. Establishing a link between regulatory focus theory and SB in older adults could potentially provide a better understanding of the motivational driving factors that lead older adults to engage more in SB.

Therefore, the aim of this study was to determine the association between prevention and promotion focus with SB in older adults and to compare these associations with two factors that are commonly known to be associated with SB among older adults (ie, age and BMI. We hypothesized that SB would have a significant positive association with age, BMI, and prevention focus, but a nonsignificant association with promotion focus.

Methods

Study Design

This cross-sectional investigation was part of a larger study that is federally funded by the National Institute on Minority Health and Health Disparities (R01MD018025) and has been pre-registered on ClinicalTrials.gov (NCT05778604). The protocol for the study has been previously published elsewhere [16]. The study was conducted in the Greater Orlando, metropolitan area, FL and recruitment was accomplished through word-of-mouth, flyer distribution, and partnership with local communities. A total of 141 community-dwelling older adults were recruited. Participants were included in this study if they were at least 60 years of age, could stand on their own, had low-income status based on the 2019 poverty thresholds relative to family size [17], had fully completed the Regulatory Focus Questionnaire (RFQ), and had at least 4 days of valid recorded ActiGraph data (each day≥10 hours of wear time). Participants who were unable to perform PA, receiving treatment from a rehabilitation facility, or hospitalized more than 3 times in the past 12 months were excluded. After screening for inclusion criteria, 93 participants were included and analyzed in this study.

Data Sources

Demographic Measurements

Demographic characteristics such as age and sex were collected using a self-reported survey. Height and weight were assessed without shoes using a digital physician scale with a built-in stadiometer (Health-O-Meter, Model 402KL, McCook). BMI was calculated using height and weight as kg/m^2.

Regulatory Focus Questionnaire

Participants were asked to complete the RFQ, an 11-item questionnaire that helps determine whether an individual's motivational orientation is primarily prevention- or promotion-focused [18,19]. The RFQ produces independent prevention and promotion focus scores. Six questions contribute to the calculation of the promotion focus score, while five contribute to the prevention focus score. Seven questions are reverse scored (three for promotion and four for prevention). The RFQ uses a 5-point Likert scale based on how frequently specific events occur or have occurred in an individual's life, ranging from 1 ("never or seldom"/ "never true"/ "certainly false") to 5 ("very often"/ "very often true"/ "certainly true"). An example of a promotion question on the RFQ is as follows: "Compared to most people, are you typically unable to get what you want out of life?" The type of motivational focus most dominantly exhibited by an individual is based on the higher of the two scores (ie, prevention or promotion). For the prevention domain, a higher score indicates a greater individual focus on prevention (eg, individuals focus on avoiding negative outcomes and ensuring safety when making decisions). Similarly for the promotion domain, a higher score reflects a greater findividual focus on promotion (eg, maximizing potential gains and seeking growth and advancement). The RFQ has internal reliability coefficients of α=0.73 for the promotion scale and α=0.80 for the prevention scale [18].

Sedentary Behavior

To objectively measure SB, an ActiGraph GT9X Link wireless activity monitor (ActiGraph LLC) was used. The GT9X Link is a small and lightweight device (3.5×3.5×1 cm; 14g) that contains a triaxial accelerometer, has a dynamic range of ±8 gravitational units (g), and can be worn on either the hip or wrist. This device was initialized to record at a sampling rate of 30Hz. Participants were instructed to wear the device on their nondominant wrist for 7 consecutive days and to remove it only

XSL•FO
RenderX

during showering or bathing. Accelerometer data were extracted from the GT9X Link using the ActiLife (version 6.13.5; ActiGraph LLC), R statistical software (version 4.3.1; R Foundation for Statistical Computing), and existing accelerometer data analysis code that use the *GGIR* R-package (version 3.1 - 0) [20-23].

Total recorded SB (in minutes) was determined using the Euclidean Norm Minus One acceleration scalar metric in milli-gravitational units (mg) and classified as <30 mg [20]. Total SB was calculated for this study by summing the amount of SB (in minutes) for all valid days for each participant. For this investigation, participants were included if they had a minimum of 4 recorded days, with a minimum of 10 hours (600 minutes) each day, and a maximum of 7 recorded days. These requirements were consistent with previous PA research using the ActiGraph GT9X Link [24]. Data filtering was performed using the MATLAB multiparadigm program.

Statistical Analyses

Minitab (version 21.4.3; Minitab LLC) and R software (version 4.3.1; R Foundation for Statistical Computing) were used to perform statistical analyses for this study. Each factor was checked for normality before analysis using descriptive statistics and Anderson-Darling normality tests through Minitab. Spearman rank correlation coefficients (ρ) were used due to the nonnormal distribution of each variable. R software was used to calculate correlation coefficients, compare statistically significant correlation values using Fisher Z transformation, and perform power analysis using the *pwr* package (version 1.3 - 0) [25]. The threshold for statistical significance was set at $P<.05$.

Ethical Considerations

This study was approved by the Institutional Review Board of the University of Central Florida (STUDY00003206), preregistered on ClinicalTrials.gov (NCT05778604), and carried out in accordance with the Declaration of Helsinki. The protocol has also been previously published elsewhere [16]. All participants gave written informed consent prior to participation. The data for this study were stored and managed using Research Electronic Data Capture (REDCap), a secure, web-based application that is designed to store data for research studies [26,27], and all participant data were deidentified. Participants were compensated US$50 in the form of a gift card upon completion of the study.

Results

After screening for inclusion criteria, a total of 93 participants were included in this analysis, with the majority being female participants (83/93, 89%). Further self-reported demographic information regarding participants (eg, education level and general health) is shown in Table 1. Sixty-seven percent (62/93) of participants had a dominant prevention focus (as indicated by a greater prevention score), while 33% (31/93) of participants had a dominant promotion focus. Descriptive statistics for each factor are presented in Table 2.

XSL•FO
RenderX

Table . Participant demographics of community-dwelling older adults.

Variables	Participants (N=93)
Age (years), mean (SD), range	75.44 (6.71), 61.76 - 89.46
Sex, n (%)	
Male	10 (11)
Female	83 (89)
Education, n (%)	
Lower than high school	14 (15)
High school	52 (56)
College or above	27 (29)
General health, n (%)	
Poor	1 (1)
Fair	15 (17)
Good	53 (57)
Very good	21 (23)
Excellent	2 (2)
Living status, n (%)	
Alone	65 (70)
With partner/spouse	13 (14)
With family/ friend	13 (14)
Other	2 (2)
Financial status, n (%)	
Much less than adequate	7 (8)
Less than adequate	20 (22)
Just enough	54 (58)
More than enough	11 (11)
Much more than enough	1 (1)

Table . Descriptive statistics for age, BMI, RFQ[a] scores (prevention and promotion scales), and total SB[b] (prevention and promotion groups) of community-dwelling older adults[c].

Variables	Median (IQR)	Min	Max	*P* value[d]
Age (years)	74 (10)	61	89	.24
BMI (kg/m^2)	29.77 (8.22)	20.95	47.03	.02
RFQ[a] Prevention Focus Score	3.80 (1.20)	1.40	5.00	.01
RFQ Promotion Focus Score	3.33 (0.83)	1.67	4.67	.005
Total SB[b] of Prevention Focus Group (min)	4989 (949)	2605	7713	.25
Total SB of Promotion Focus Group (min)	4851 (981)	2527	7660	.03

[a]RFQ: Regulatory Focus Questionnaire.

[b]SB: sedentary behavior.

[c] Due to having a normal distribution, the mean and standard deviation for the age factor (as opposed to the median and IQR) are presented elsewhere.

[d]*P* value represents the results from the Anderson-Darling normality tests ($P<.05$ indicates nonnormal distribution).

XSL•FO
RenderX

Correlational data is presented in Table 3 . The prevention focus score from the RFQ had the highest statistically significant correlation with total SB among the considered factors (ρ=0.296; *P*=.004). BMI had the second highest statistically significant correlation with total SB (ρ=0.204, *P*=.049). It was found through Fisher Z transformation that the correlation between prevention focus and SB was not significantly higher than that between BMI and SB (*P*=.51). Overall, the significant associations were found to be relatively weak. The association between RFQ promotion focus score and total SB was nonsignificant and near zero (ρ=0.002, *P*=.99). Age also had a weak, nonsignificant association with total SB (ρ=0.116, *P*=.27). Power analyses indicated that only the association between prevention focus and SB achieved adequate power of at least 80% (82.86%), exceeding the required sample size needed for achieving the Spearman coefficient of 0.296 (N=84). All other associations required a sample size that exceeded the total study sample size of 93 participants to achieve adequate power and therefore had relatively lower and inadequate power (<50%).

Table . Associations between age, BMI, RFQ[a] prevention score, and RFQ promotion score with SB[b].

Variables	Spearman rank correlation (ρ)	*P* value
Age	0.116	.27
BMI	0.204	.049
RFQ[a] prevention focus score	0.296	.004
RFQ promotion focus score	0.002	.99

[a]RFQ: Regulatory Focus Questionnaire.

[b]SB: sedentary behavior.

Discussion

Principal Findings

The purpose of this study was to determine the association between prevention and promotion focus with SB in older adults, as well as to compare those associations with two factors that are commonly known to have an association with SB among older adults(ie, age and BMI). Analysis showed that SB has a significant positive association with BMI and prevention focus. This finding supports our initial hypothesis. Consistent with the hypothesis, promotion focus was not significantly associated with SB. Contrary to expectations, we did not observe an association between age and SB. BMI was associated with SB, but this association with SB was relatively weaker than that of prevention focus. Overall, the associations were weaker than expected. Comparative analysis of the significant correlation values using Fisher Z transformation showed that the correlation between prevention focus and SB was not significantly higher than the correlation between BMI and SB.

A key finding of this study is the observation of a weak, but positive association between a prevention-focused regulatory style and SB, indicating that individuals with a primarily prevention-focused mindset may engage in increased amounts of SB. This result may be attributed to the cautious, safe, and loss-avoidant mindset associated with the prevention focus orientation. This result aligns with previous work, which found that safety concerns of older adults are a barrier to reducing their sedentary time, whether in their own homes or in their neighborhoods [28,29]. Individuals that exhibit such personality traits may be more likely to avoid engaging in PA and overall movement due to the possibility of injury. An injury may result in losses, specifically the possible expenses associated with recovery, as well as the loss of time needed for physical recovery and possible loss of independence. Individuals with such mindsets may view PA or increased movement as dangerous or unsafe, thus leading to avoidance. Environmental factors such as a neighborhood safety can also contribute to this mindset, further preventing PA and movement outside of one's home if an individual perceives that their safety in their neighborhood could be at risk.

The promotion focus scores and total SB were weakly associated, which may have been attributed to the possibility that decreased SB, as well as the benefits associated with it, are not considered as growth or advancement to older adults. Furthermore, older adults' decision to engage in SB may be more influenced by a prevention focus because the desires associated with it (eg, desires for safety, avoidance of negative consequences, and avoidance of losses) are perceived as more of a "need" rather than a desire. In contrast, the desires associated with a promotion motivation orientation (eg, desire to achieve growth, advancement, and maximize potential gains), which could be considered as less of a "need."

It was expected that BMI would have a significant positive association with SB, due to the directly proportional association between BMI and weight. Increased time in SB can potentially lead to weight gain and has been positively related to the risk of obesity and diabetes [30-33]. Furthermore, studies have shown that SB is positively associated with BMI [34,35]. Interestingly, despite the difference being statistically insignificant, the prevention focus scores from the RFQ had a relatively greater correlation coefficient value than BMI when compared to total SB. For years, BMI has been used as an index of obesity to classify individuals based on their body weight relative to height. BMI continues to be used in medical practice and research due to being an affordable and readily obtained metric [36]. However, BMI has faced numerous backlashes, primarily due to its inability to differentiate between body lean mass and body fat mass [36-39]. For example, an individual with a greater amount of body lean mass could have the same BMI as an individual with a greater amount of body fat mass. This variability present in BMI could explain the relatively lower correlation value when relating to SB. Furthermore, the

XSL•FO
RenderX

motivation orientation of an individual, specifically an older adult can influence their decision-making and behavior, which could drive them to engage in SB, therefore directly resulting in changes to their BMI.

Various studies have shown a positive correlation between age and SB in older adults [40,41]. However, a nonsignificant association between the two variables was observed in this study. Previous work has found that a multitude of factors contribute to the SB levels of older adults beyond age, such as accessibility to transportation, living environment, and weather [42]. The sites from which participants in this study were recruited from, varied in terms of social life. Some sites held weekly social activities (eg, exercise classes and bingo nights), while others did not hold any social events. These differences could adversely affect the decision of older adults to engage in SB and could explain the weak and nonsignificant association between age and total SB in this study.

Strengths, Limitations, and Future Directions

Overall, despite the weak associations, the findings of this study show that the characteristics associated with a prevention focus may be a factor that leads older adults to engage in SB. The results from this study support the idea that older adults might be less inclined to engage in SB if they are given suggestions on how to maintain safety during PA or reassured that it is acceptable to make mistakes. This idea can be used to create new or improve existing intervention programs aimed toward improving the PA levels of older adults, which can ultimately result in a reduction in their fall risk. Spearman rank correlation was chosen for the analysis of this study due to its ability to adequately manage both normally and nonnormally distributed data. Although an appropriate test was used to perform correlation analysis, certain limitations exist within this study. First, a power analysis showed that only the association between the prevention focus score from the RFQ and total SB produced adequate power (82.86%). Although this value indicates reasonable sensitivity for detecting the observed association, all other associations produced a power significantly below this value (<50%).

Second, this study did not identify the causal relationship between the considered factors and only focused on pairwise correlations, as opposed to considering confounding effects. For example, there was a significantly greater number of females included in the sample than male participants (83/93). It has been previously found that female participants tend to be more prevention-focused than male participants [43], which explains the larger number of prevention-focused individuals (62/93). The exploration of confounding effects could provide a deeper understanding of the associations presented in this study. Future work should involve increasing the sample size to include more promotion-focused individuals and raising the power of the correlation tests to adequately determine the associations between the factors, as well as using techniques such as regression to incorporate confounders and conduct formal causal analysis.

Conclusions

Although many environmental factors drive older adults to engage in SB, psychological factors also play a role in influencing this behavior. Prevention focus, as determined by the RFQ was found to have a weak but significant positive association with SB in older adults. The association between BMI and SB was relatively weaker in comparison, but no significant differences were found between the two associations. These findings suggest that older adults may be driven to engage in increased amounts of SB due to having a dominant prevention focus, which revolves around thoughts of safety and the avoidance of negative consequences. Identifying an older adult's motivational orientation can provide useful insight into understanding their decisions to engage in SB and further identify what strategic approach would be most effective when setting goals and implementing safeguards to decrease SB.

Acknowledgments

This work was supported by the National Institute on Minority Health and Health Disparities of the National Institutes of Health under award R01MD018025 and the Office of the Director, Chief Officer for Scientific Workforce Diversity of the National Institutes of Health under award number 3R01MD018025-02S1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The organizations mentioned had no roles in the design of this study.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JRS, AB, KL, JP
Funding acquisition: LT
Investigation: JRS, AB, KL, JHP, LT
Methodology: JHP, RX, NL, LT
Project administration: LT
Supervision: JHP, LT
Writing – original draft: JRS
Writing – review & editing: JRS, AB, KL, JHP, RX, NL, LT

XSL•FO
RenderX

Conflicts of Interest

JHP, RX, NL, and LT received financial support from the National Institute on Minority Health and Health Disparities of the National Institutes of Health under award R01MD018025. JRS, AB, KL, and LT received financial support from the Office of the Director, Chief Officer for Scientific Workforce Diversity of the National Institutes of Health under award number 3R01MD018025-02S1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Chastin SFM, Buck C, Freiberger E, et al. Systematic literature review of determinants of sedentary behaviour in older adults: a DEDIPAC study. Int J Behav Nutr Phys Act 2015 Oct 6;12(127):127. [doi: 10.1186/s12966-015-0292-3] [Medline: 26437960]
2. Pesola AJ, Laukkanen A, Tikkanen O, Finni T. Heterogeneity of muscle activity during sedentary behavior. Appl Physiol Nutr Metab 2016 Nov;41(11):1155-1162. [doi: 10.1139/apnm-2016-0170] [Medline: 27806635]
3. Fountaine CJ, Johann J, Skalko C, Liguori GA. Metabolic and energy cost of sitting, standing, and a novel sitting/stepping protocol in recreationally active college students. Int J Exerc Sci 2016;9(2):223-229. [doi: 10.70252/WHAR7442] [Medline: 27182423]
4. 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]
5. Chastin S, Gardiner PA, Harvey JA, et al. Interventions for reducing sedentary behaviour in community-dwelling older adults. Cochrane Database Syst Rev 2021 Jun 25;6(6):CD012784. [doi: 10.1002/14651858.CD012784.pub2] [Medline: 34169503]
6. King AC, Hekler EB, Grieco LA, et al. Harnessing different motivational frames via mobile phones to promote daily physical activity and reduce sedentary behavior in aging adults. PLoS ONE 2013;8(4):e62613. [doi: 10.1371/journal.pone.0062613] [Medline: 23638127]
7. Lewis LK, Rowlands AV, Gardiner PA, Standage M, English C, Olds T. Small Steps: preliminary effectiveness and feasibility of an incremental goal-setting intervention to reduce sitting time in older adults. Maturitas 2016 Mar;85:64-70. [doi: 10.1016/j.maturitas.2015.12.014] [Medline: 26857881]
8. Kerr J, Takemoto M, Bolling K, et al. Two-arm randomized pilot intervention trial to decrease sitting time and increase sit-to-stand transitions in working and non-working older adults. PLoS ONE 2016;11(1):e0145427. [doi: 10.1371/journal.pone.0145427] [Medline: 26735919]
9. Rietzschel EF, Ritter SM. Chapter 1 - moving from creativity to innovation. In: Reiter-Palmon R, Kennel VL, Kaufman JC, editors. Individual Creativity in the Workplace: Academic Press; 2018:3-34. [doi: 10.1016/B978-0-12-813238-8.00001-2]
10. Higgins ET. Beyond pleasure and pain. Am Psychol 1997 Dec;52(12):1280-1300. [doi: 10.1037//0003-066x.52.12.1280] [Medline: 9414606]
11. Higgins ET. Beyond Pleasure and Pain: How Motivation Works: Oxford University Press; 2011.
12. Fawsitt F, Dockray S, Setti A. Regulatory focus and perceptions of ageing: exploring the connections. Aging Ment Health 2022 Jul;26(7):1451-1459. [doi: 10.1080/13607863.2021.1952550] [Medline: 34282679]
13. Levy BR, Slade MD, Kasl SV. Longitudinal benefit of positive self-perceptions of aging on functional health. J Gerontol B Psychol Sci Soc Sci 2002 Sep;57(5):409-417. [doi: 10.1093/geronb/57.5.p409] [Medline: 12198099]
14. Kotter-Grühn D, Kleinspehn-Ammerlahn A, Gerstorf D, Smith J. Self-perceptions of aging predict mortality and change with approaching death: 16-year longitudinal results from the Berlin Aging Study. Psychol Aging 2009 Sep;24(3):654-667. [doi: 10.1037/a0016510] [Medline: 19739922]
15. Wurm S, Tomasik MJ, Tesch-Römer C. Serious health events and their impact on changes in subjective health and life satisfaction: the role of age and a positive view on ageing. Eur J Ageing 2008 Jun;5(2):117-127. [doi: 10.1007/s10433-008-0077-5] [Medline: 28798566]
16. Thiamwong L, Xie R, Park JH, Lighthall N, Loerzel V, Stout J. Optimizing a technology-based body and mind intervention to prevent falls and reduce health disparities in low-income populations: protocol for a clustered randomized controlled trial. JMIR Res Protoc 2023 Oct 3;12:e51899. [doi: 10.2196/51899] [Medline: 37788049]
17. Poverty thresholds.: United States Census Bureau; 2020. URL: https://www.census.gov/data/tables/time-series/demo/income-poverty/historical-poverty-thresholds.html [accessed 2024-06-02]
18. Higgins ET, Friedman RS, Harlow RE, Idson LC, Ayduk ON, Taylor A. Achievement orientations from subjective histories of success: promotion pride versus prevention pride. Euro J Social Psych 2001 Jan;31(1):3-23. [doi: 10.1002/ejsp.27]
19. Hodis FA, Hodis GM. Assessing motivation of secondary school students: an analysis of promotion and prevention orientations as measured by the Regulatory Focus Questionnaire. J Psychoeduc Assess 2017 Oct;35(7):670-682. [doi: 10.1177/0734282916658385]
20. Accelerometer data processing with GGIR. The comprehensive R archive network.(CRAN). URL: https://cran.r-project.org/web/packages/GGIR/vignettes/GGIR.html [accessed 2024-05-01]

XSL•FO
RenderX

21. Sabia S, van Hees VT, Shipley MJ, et al. Association between questionnaire- and accelerometer-assessed physical activity: the role of sociodemographic factors. Am J Epidemiol 2014 Mar 15;179(6):781-790. [doi: 10.1093/aje/kwt330] [Medline: 24500862]
22. Migueles JH, Rowlands AV, Huber F, Sabia S, van Hees VT. Ggir: A research community–driven open source r package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. Journal for the Measurement of Physical Behaviour 2019;2(3):188-196. [doi: 10.1123/jmpb.2018-0063]
23. van Hees VT, Gorzelniak L, Dean León EC, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS ONE 2013;8(4):e61691. [doi: 10.1371/journal.pone.0061691] [Medline: 23626718]
24. Gorman E, Hanson HM, Yang PH, Khan KM, Liu-Ambrose T, Ashe MC. Accelerometry analysis of physical activity and sedentary behavior in older adults: a systematic review and data analysis. Eur Rev Aging Phys Act 2014;11(1):35-49. [doi: 10.1007/s11556-013-0132-x] [Medline: 24765212]
25. Pwr: basic functions for power analysis. The comprehensive R archive network CRAN. 1988. URL: https://cran.r-project.org/package=pwr [accessed 2024-05-01]
26. 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]
27. 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]
28. Fanning J, Nicklas BJ, Rejeski WJ. Intervening on physical activity and sedentary behavior in older adults. Exp Gerontol 2022 Jan;157(111634):111634. [doi: 10.1016/j.exger.2021.111634] [Medline: 34826574]
29. Voss ML, Pope JP, Copeland JL. Reducing sedentary time among older adults in assisted living: perceptions, barriers, and motivators. Int J Environ Res Public Health 2020 Jan 22;17(3):717. [doi: 10.3390/ijerph17030717] [Medline: 31979131]
30. Hu FB, Li TY, Colditz GA, Willett WC, Manson JE. Television watching and other sedentary behaviors in relation to risk of obesity and type 2 diabetes mellitus in women. JAMA 2003 Apr 9;289(14):1785-1791. [doi: 10.1001/jama.289.14.1785] [Medline: 12684356]
31. Gómez-Cabello A, Vicente-Rodríguez G, Pindado M, et al. Increased risk of obesity and central obesity in sedentary postmenopausal women. Nutr Hosp 2012;27(3):865-870. [doi: 10.3305/nh.2012.27.3.5668] [Medline: 23114948]
32. Kronenberg F, Pereira MA, Schmitz MK, et al. Influence of leisure time physical activity and television watching on atherosclerosis risk factors in the NHLBI Family Heart Study. Atherosclerosis 2000 Dec;153(2):433-443. [doi: 10.1016/s0021-9150(00)00426-3] [Medline: 11164433]
33. Jakes RW, Day NE, Khaw KT, et al. Television viewing and low participation in vigorous recreation are independently associated with obesity and markers of cardiovascular disease risk: EPIC-Norfolk population-based study. Eur J Clin Nutr 2003 Sep;57(9):1089-1096. [doi: 10.1038/sj.ejcn.1601648] [Medline: 12947427]
34. Stamatakis E, Davis M, Stathi A, Hamer M. Associations between multiple indicators of objectively-measured and self-reported sedentary behaviour and cardiometabolic risk in older adults. Prev Med 2012 Jan;54(1):82-87. [doi: 10.1016/j.ypmed.2011.10.009] [Medline: 22057055]
35. Gennuso KP, Gangnon RE, Matthews CE, Thraen-Borowski KM, Colbert LH. Sedentary behavior, physical activity, and markers of health in older adults. Med Sci Sports Exerc 2013 Aug;45(8):1493-1500. [doi: 10.1249/MSS.0b013e318288a1e5] [Medline: 23475142]
36. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. Nutr Today 2015 May;50(3):117-128. [doi: 10.1097/NT.0000000000000092] [Medline: 27340299]
37. Flegal KM, Shepherd JA, Looker AC, et al. Comparisons of percentage body fat, body mass index, waist circumference, and waist-stature ratio in adults. Am J Clin Nutr 2009 Feb;89(2):500-508. [doi: 10.3945/ajcn.2008.26847] [Medline: 19116329]
38. Romero-Corral A, Somers VK, Sierra-Johnson J, et al. Normal weight obesity: a risk factor for cardiometabolic dysregulation and cardiovascular mortality. Eur Heart J 2010 Mar;31(6):737-746. [doi: 10.1093/eurheartj/ehp487] [Medline: 19933515]
39. Wellens RI, Roche AF, Khamis HJ, Jackson AS, Pollock ML, Siervogel RM. Relationships between the body mass index and body composition. Obes Res 1996 Jan;4(1):35-44. [doi: 10.1002/j.1550-8528.1996.tb00510.x] [Medline: 8787936]
40. Lord S, Chastin SFM, McInnes L, Little L, Briggs P, Rochester L. Exploring patterns of daily physical and sedentary behaviour in community-dwelling older adults. Age Ageing 2011 Mar;40(2):205-210. [doi: 10.1093/ageing/afq166] [Medline: 21239410]
41. Arnardottir NY, Koster A, Van Domelen DR, et al. Objective measurements of daily physical activity patterns and sedentary behaviour in older adults: Age, Gene/Environment Susceptibility-Reykjavik Study. Age Ageing 2013 Mar;42(2):222-229. [doi: 10.1093/ageing/afs160] [Medline: 23117467]
42. Tam-Seto L, Weir P, Dogra S. Factors influencing sedentary behaviour in older adults: an ecological approach. AIMS Public Health 2016;3(3):555-572. [doi: 10.3934/publichealth.2016.3.555] [Medline: 29546182]
43. Gutermuth D, Hamstra MRW. Are there gender differences in promotion-prevention self-regulatory focus? Br J Psychol 2024 May;115(2):306-323. [doi: 10.1111/bjop.12688] [Medline: 37984379]

XSL•FO
RenderX

Abbreviations

PA: physical activity
REDCap: Research Electronic Data Capture
RFQ: Regulatory Focus Questionnaire
SB: sedentary behavior

Edited by TDA Cardoso; submitted 14.06.24; peer-reviewed by JE Teixeira; revised version received 11.03.25; accepted 12.03.25; published 01.05.25.

Please cite as:
Suarez JR, Blount A, Lafontant K, Park JH, Xie R, Lighthall N, Thiamwong L
Association Between Prevention Focus and Sedentary Behavior in Older Adults: Cross-Sectional Study
Interact J Med Res 2025;14:e63280
URL: https://www.i-jmr.org/2025/1/e63280
doi:10.2196/63280

© Jethro Raphael Suarez, Amber Blount, Kworweinski Lafontant, Joon-Hyuk Park, Rui Xie, Nichole Lighthall, Ladda Thiamwong. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 1.5.2025. 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

Associations Between Sleep Duration and Activity of Daily Living Disability Among Older Adults in China: Cross-Sectional Study

Huimin Fan[1], PhD; Weijie Yu[2], BMed; Hongguo Rong[3*], PhD; Xiaokun Geng[1,4,5*], MD, PhD

[1]Department of Neurology, Beijing Luhe Hospital, Capital Medical University, Beijing, China

[2]School of Traditional Chinese Medicine, Beijing University of Chinese Medicine, Beijing, China

[3]Institute for Excellence in Evidence-Based Chinese Medicine, Beijing University of Chinese Medicine, No. 11 Beisanhuan East Road, Chaoyang District, Beijing, China

[4]China-America Institute of Neuroscience, Beijing Luhe Hospital, Capital Medical University, Beijing, China

[5]Stroke Center, Department of Neurology, Beijing Luhe Hospital, Capital Medical University, No. 82 Xinhua SouthRoad, Tongzhou District, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Xiaokun Geng, MD, PhD
Department of Neurology, Beijing Luhe Hospital, Capital Medical University, Beijing, China

Abstract

Background: China has the largest elderly population globally; the growth rate of the aged tendency of the population was higher than that of Western countries. Given the distinctions in historical, ethnic, and economic status as well as socio-cultural background, Chinese adults had different sleep patterns compared with adults in other countries. Considering the heavy disease burden caused by activities of daily living (ADL) disability, we conducted a cross-sectional analysis using data from the China Health and Retirement Longitudinal Study (CHARLS) to test the hypothesis that individuals with short and longer sleep duration are more likely to have ADL disability.

Objective: ADL disability is a common condition affecting the quality of life among older people. This study aimed to explore the associations between sleep duration and ADL disability among middle-aged and older adults in China.

Methods: This cross-sectional study used data from 17,607 participants from the 2018 CHARLS (from 2018 to 2020), an ongoing representative survey of adults aged 45 years or older and their spouses. Self-reported sleep duration per night was obtained from face-to-face interviews. The ADL was measured using a 6-item summary assessed with an ADL scale that included eating, dressing, getting into or out of bed, bathing, using the toilet, and continence. Multiple generalized linear regression models—adjusted for age, sex, education, marital status, tobacco and alcohol use, depression, place of residence, sensory impairment, self-reported health status, life satisfaction, daytime napping, chronic disease condition, and sample weights—were used.

Results: Data were analyzed from 17,607 participants, of whom 8375 (47.6%) were men. The mean (SD) age was 62.7 (10.0) years. Individuals with 4 hours or less (odds ratio [OR] 1.91, 95% CI 1.60 - 2.27; *P*<.001), 5 hours (OR 1.33, 95% CI 1.09 - 1.62; *P*=.006), 9 hours (OR 1.48, 95% CI 1.13 - 1.93; *P*<.001), and 10 hours or more (OR 1.88, 95% CI 1.47 - 2.14; *P*<.001) of sleep per night had a higher risk of ADL disability than those in the reference group (7 hours per night) after adjusting for several covariates. Restricted cubic splines analysis suggested a U-shaped association between sleep duration and ADL disability. When sleep duration fell below 7 hours, an increased sleep duration was associated with a significantly low risk of ADL disability, which was negatively correlated with sleep duration until it fell below 7 hours (OR 0.83, 95% CI 0.79 - 0.87; *P*<.001). When sleep duration exceeded 7 hours, the risk of ADL disability would increase facing prolonged sleep duration (OR 1.19, 95% CI 1.12 - 1.27; *P*<.001). ADL disability should be monitored in individuals with insufficient (≤4 or 5 hours per night) or excessive (9 or ≥10 hours per night) sleep duration.

Conclusions:: In this study, a U-shaped association between sleep duration and ADL disability was found. Future longitudinal studies are needed to establish temporality and examine the mechanisms of the associations between sleep duration and ADL disability.

(Interact J Med Res 2025;14:e65075) doi:10.2196/65075

XSL•FO
RenderX

KEYWORDS

sleep; sleep duration; activities of daily living; CHARLS; survey; questionnaire; self-reported; gerontology; geriatric; older adult; elder; elderly; aging; ADL; physical function; physical functioning; well-being; association; correlation; China Health and Retirement Longitudinal Study

Introduction

The World Health Organization (WHO) survey of 2022 reported that 46.1% of adults aged 60 years and older are living with a disability globally [1]. Furthermore, according to the WHO projections, there will be 66 million older adults in China who live with a functional handicap by 2050 [1]. Activities of daily living (ADL) disability are one of the most common health problems in older adults, affecting about 1 in 6 older adults worldwide [2]. The prevalence of ADL disability increased with aging and had become a significant factor that increases the risk of mortality in older adults [3,4]. ADL disability seriously affected the quality of life and placed a high burden on care providers and the care system. Current research has focused on the prevention of, rehabilitation from, or management of ADL disability, with growing interest in the role of sleep patterns, such as sleep duration [5-7].

Sleep duration has become a significant public health issue [8], associated with the increasing risk of cardiovascular diseases, stroke, and mortality [9-11]. The WHO launched the ”World Mental Health Report—Transforming Health for All” on June 17, 2022 [12]. This report emphasizes that sleep loss is an extremely important risk factor for mental health issues. At present, the relationship between sleep duration and ADL disability has received much attention [13]. A longitudinal observational study suggested that short sleep duration during inpatient rehabilitation might be against greater functional ability at discharge among individuals with acute stroke [14]. The data from the National Health and Nutrition Examination Survey in the United States found that only short sleep duration was significantly associated with ADL disability [15]. The New Integrated Suburban Seniority Investigation Project study also reported that sleeping <6 hours/day was associated with a higher risk of incident disability among older Japanese people [16]. Although the above study did not find a relationship between long sleep and physical activity, the data from the English Longitudinal Study of Ageing cohort showed that both reduced and increased sleep duration were associated with limited agility and mobility disability [17]. A cross-sectional study in Taiwan also reported long sleep duration had a positive association with grip strength performance among older adults [18]. There is still no unified conclusion in the research on the relationship between sleep duration and ADL disability. So far, research on the relationship between sleep duration and ADL disability is well worth conducting.

China has the largest population of older adults globally; the growth rate of the aged tendency of the population was higher than that of Western countries [19,20]. Given the distinctions in historical, ethnic, and economic status as well as socio-cultural backgrounds, Chinese adults had different sleep patterns compared to those from other countries [21,22]. Considering the heavy disease burden caused by ADL disability, we conducted a cross-sectional analysis using data from the China Health and Retirement Longitudinal Study (CHARLS) to test the hypothesis that individuals with short and longer sleep duration are more likely to have ADL disability.

Methods

Ethical Considerations

The CHARLS protocol was approved by the Peking University Biomedical Ethics Review Board (IRB00001052-11015) [23,24]. This cross-sectional study adhered to the reporting guidelines stipulated by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline [25]. According to the Common Rule (45 CFR part 46), this study was exempt from institutional review board approval and the requirement for informed patient consent because we did not use clinical data or involve human participants.

Study Participants

This cross-sectional study used data from the 2018 survey of the CHARLS conducted from 2018 to 2020, a survey of residents aged 45 years or older in 28 provinces across China [26,27]. The national baseline survey of the CHARLS included 17,708 respondents in 150 counties or districts and 450 villages or urban communities throughout the country. Detailed methods of the CHARLS have been described previously [23]. We applied the latest CHARLS wave in 2018 to 19,816 individuals to investigate the relationship between sleep duration and ADL disability. In the current study, we formulated the inclusion criteria for the study subjects: age ≥45 years, respondents with complete information on sleep duration, the ADL scale, and other covariates. We excluded the following outliers: missing date of sample weights, birth date, sleep duration, and ADL scale.

Sleep Duration

Similar to the English Longitudinal Study of Aging [28] in England, the CHARLS evaluated subjects’ sleep duration by utilizing the questions: “During the past month, how many hours of actual sleep did you get at night (average hours for one night)? (This might be shorter than the number of hours you spend in bed).” The response was an average sleep duration of typical usual weekdays or workdays within a month. Consistent with the previous CHARLS sleep study [29-32], we separated respondents into 7 sleep duration groups (≤4, 5, 6, 7, 8, 9, and ≥10 hours out of each evening) in the research. The CHARLS used this question to assess participants’ daytime napping: “In the past month, how long have you taken on average for a nap after lunch?” On the basis of 4 nap time groups considering the results [33,34], the participants were classified as nonnappers (0 min), short nappers (<30 minutes), moderate nappers (30 - 90 minutes), and extended nappers (>90 minutes).

XSL•FO
RenderX

ADL Disability

The CHARLS employed a 6-item summary derived from an ADL scale to quantify ADL utilizing a comprehensive series of inquiries that asked: "Are you experiencing any difficulties with performing any everyday activity due to physical, mental, emotional, or memory impairments, excluding any that you anticipate will persist for less than three months?" The daily tasks encompassed dressing, bathing or showering, eating, getting into or out of bed, using the toilet, and controlling urination and defecation. The response scale contained 4 options: (1) No, I don't have any difficulty; (2) I have difficulty but can still do it; (3) Yes, I have difficulty and need help; and (4) I cannot do it. The ADL scale was extensively used in previous studies among Chinese older adults and has shown good reliability and validity [35,36]. Individuals classified as independent possessed the capability to perform all 6 activities without any difficulty, whereas those who indicated a need for assistance in any of the aforementioned tasks were considered to have an ADL disability [37,38].

Covariates

The CHARLS collected information on sociodemographic characteristics and health-related factors using a structured questionnaire. Covariates that might confound the associations included age, sex, education (elementary school and below, secondary school, and college and above), marital status (married and others), residence (rural or urban), smoking and drinking status (never vs current), visual impairment, hearing impairment, self-reported general health status (ie, very good, good, fair, poor, and very poor), self-reported life satisfaction (ie, completely satisfied, very satisfied, somewhat satisfied, not very satisfied, or not at all satisfied), daytime napping, depression, chronic disease condition (ie, none, mild, and severe), and sample weights. Respondents were deemed to have visual or hearing impairment if they reported fair or poor vision (for either long-distance or near vision) or hearing difficulties. The respondents' chronic disease status was documented based on the count of self-reported noncommunicable diseases, encompassing hypertension, diabetes, dyslipidemia, heart conditions, stroke, renal diseases, asthma, pulmonary disorders, arthritis, liver ailments, and gastric issues. As for sampling weights, the CHARLS constructed cross-sectional sample weights directly from the sampling probabilities for households and individuals, taking into account death and divorce. In the 2018 survey, the CHARLS provides two sets of cross-sectional household weights, one with corrections for nonresponse and one without. Our study adopted the sets of cross-sectional individual weights with corrections.

Statistical Analysis

In the current study, comparisons of respondents' general characteristics based on ADL disability status were conducted using analysis of variance (ANOVA) for numerical variables and ordinal χ^2 tests for categorical variables. Numerical variables were presented as the mean (SD). Categorical variables were reported as numbers and percentages. Given the dichotomous nature of the ADL disability measure utilized in this study, we adopted multivariable generalized linear models with binomial family and log links to examine the associations between sleep duration and ADL disability; meanwhile, the odds ratios (ORs) and 95% CIs were reported. Restricted cubic splines with four knots at the 5th, 35th, 65th, and 95th centiles were used to examine the association between sleep duration per night and ADL disability. The generalized linear analytic models incorporated a range of covariates, including age, sex, education, marital status, tobacco and alcohol consumption, depression, place of residence, sensory impairment, self-reported health status, life satisfaction, daytime napping, chronic disease status, and sample weights. Multiple imputation with 5 replications and a chained equation approach was adopted to deal with missing data. We also conducted analyses stratified by sex to investigate sex-specific associations. A two-sided $P<.05$ was considered to be statistically significant. All data analyses were performed with Stata version 14.0 (StataCorp).

Results

Sample Characteristics

A schematic flow diagram of the study sample is shown in Figure 1. After filtering, our final study sample included 17,607 people with and without ADL disability.

XSL•FO
RenderX

Figure 1. Flowchart of the study sample from the 2018 China Health and Retirement Longitudinal Study (CHARLS).

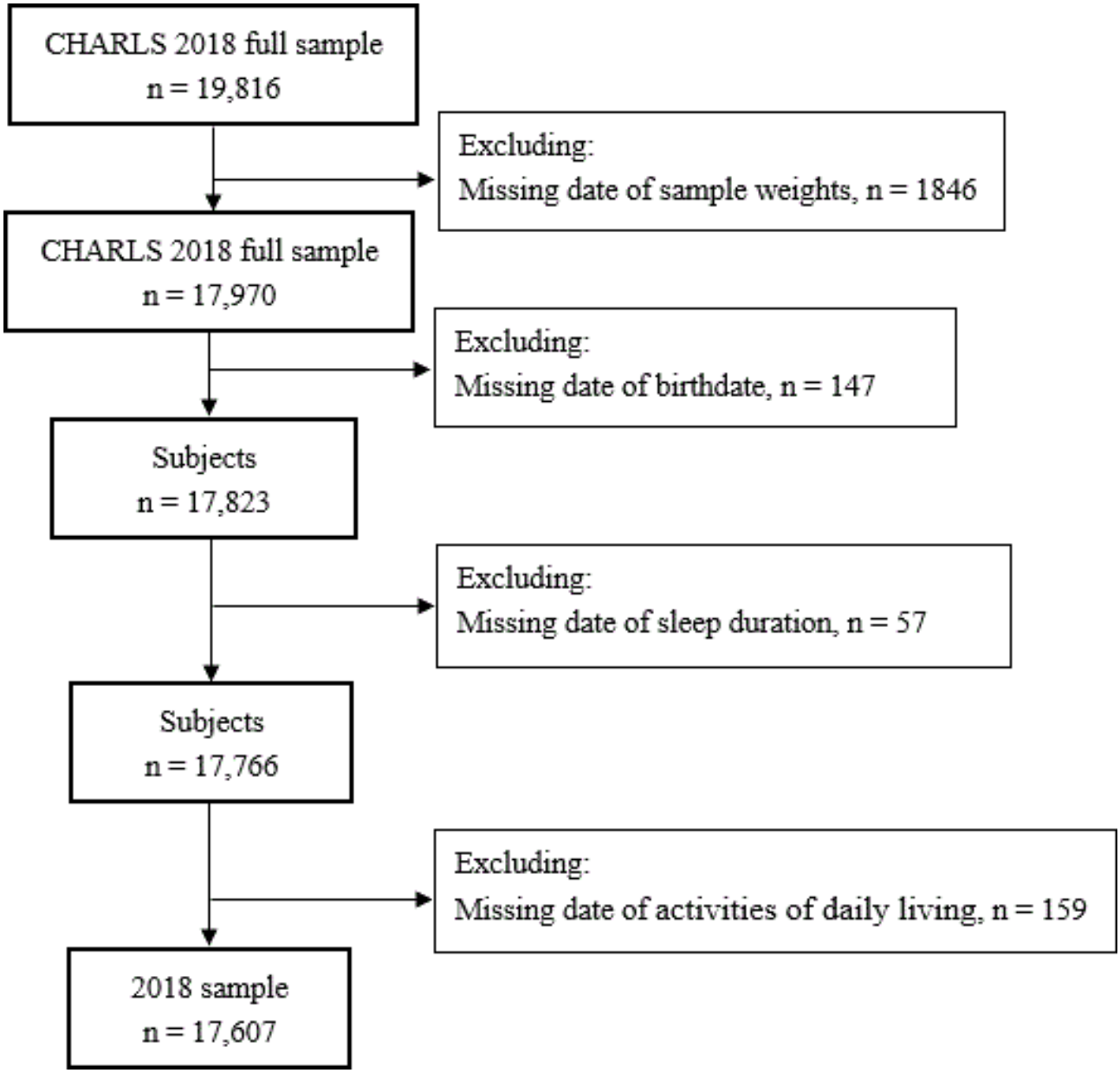


Table 1 shows the characteristics of the study population in the 2018 wave of the CHARLS. Among the 17,607 participants, 8375 (47.6%) were men and 9232 (52.4%) were women, with a mean (SD) age of 62.7 (10.0) years. A total of 3350 participants (19.0%) reported ADL disability. Compared with people with normal ADL function, individuals with ADL disability were more likely to have extremely shorter (≤4 hours) sleep duration and longer (≥10 hours) sleep duration (all P<.001). Participants with ADL disability also tended to have long napping during the day (≥90 minutes). In addition, there were significant differences between ADL disability groups concerning sex, education, tobacco and alcohol use, married status, residence, depressive status, chronic disease conditions, sensory impairments, life satisfaction, and self-reported general health status (all P<.001).

XSL•FO
RenderX

Table . Characteristics of participants by depression status in the 2018 China Health and Retirement Longitudinal Study[a].

Characteristics	Total sample (N=17,607)	ADL[b] disability (n=3350)	No ADL disability (n=14,257)	*P* value
Sleep duration per night, hours, mean (SD)	6.2 (2.0)	5.7 (2.5)	6.3 (1.9)	<.001
Sleep duration per night, hours, n (%)				<.001
≤4	3205 (18.2)	1091 (32.6)	2114 (14.8)	
5	2604 (14.8)	493 (14.7)	2111 (14.8)	
6	3847 (21.9)	552 (16.5)	3295 (23.1)	
7	2936 (16.7)	359 (10.7)	2577 (18.1)	
8	3314 (18.8)	450 (13.4)	2864 (20.1)	
9	804 (4.6)	158 (4.7)	646 (4.5)	
≥10	897 (5.1)	247 (7.4)	650 (4.6)	
Age, mean (SD), y	62.7 (10.0)	68.0 (10.2)	61.5 (9.5)	<.001
Sex, n (%)				<.001
Men	8375 (47.6）	1253 (37.4)	7122 (50.0)	
Women	9232 (52.4）	2097 (62.6)	7135 (50.1)	
Education, n (%)				<.001
Illiterate	4176 (23.7)	1250 (37.3)	2926 (20.5)	
Middle school and below	11,321 (64.3)	1913 (57.1)	9408 (66.0)	
High school and above	2110 (12.0)	187 (5.6)	1923 (13.5)	
Tobacco use, n (%)				<.001
Never	10,074 (57.8)	2063 (62.4)	8011 (56.8)	
Current	7349 (42.2)	1243 (37.6)	6106 (43.3)	
Alcohol use, n (%)				<.001
Never	11,442 (65.0)	2479 (74.0)	8963 (62.9)	
Current	6165 (35.0)	871 (26.0)	5294 (37.1)	
Married, n (%)	13,843 (78.6)	2369 (70.7)	11,474 (80.5)	<.001
Residence, n (%)				<.001
Rural	13,997 (79.7)	2812 (84.0)	11,185 (78.7)	
Urban	3567 (20.3)	534 (16.0)	3033 (21.3)	
Depression, n (%)	6182 (35.1)	1859 (55.5)	4323 (30.3)	
Daytime napping, minutes, n (%)				<.001
None	6727 (38.2)	1301 (38.8)	5426 (38.1)	
≤30	1462 (8.3)	279 (8.3)	1183 (8.3)	
31 - 90	6839 (38.8)	1204 (35.9)	5635 (39.5)	
≥90	2579 (14.7)	566 (16.9)	2013 (14.1)	
Chronic disease conditions, n (%)				<.001
None	9816 (55.8)	1368 (40.8)	8448 (59.3)	
Mild	6736 (38.3)	1595 (47.6)	5141 (36.1)	
Severe	1055 (6.0)	387 (11.6)	668 (4.7)	
Visual impairment, n (%)	14,535 (82.9)	3029 (91.5)	11,506 (80.9)	<.001

XSL•FO
RenderX

Characteristics	Total sample (N=17,607)	ADL[b] disability (n=3350)	No ADL disability (n=14,257)	*P* value
Hearing impairment, n (%)	11,706 (66.6)	2695 (80.7)	9011 (63.3)	<.001
Life satisfaction (satisfied), n (%)	14,365 (81.6)	2267 (67.7)	12,098 (84.9)	<.001
Self-reported general health status (good), n (%)	3846 (21.8)	218 (6.5)	3628 (25.5)	<.001

[a]Missing data for the following characteristics: tobacco use (184, 1.0%), residence (43, 0.2%), visual impairment (77, 0.4%), hearing impairment (35, 0.2%), life satisfaction (1354, 7.7%), and self-reported general health status (1241, 7.0%).

[b]ADL: activities of daily living.

Association Between Sleep Duration and ADL Disability

The association between sleep duration and ADL disability is shown in Table 2. In the model adjusting for age and sex (Model 1), individuals who slept ≤4 hours, 5 hours, 6 hours, 9 hours, and ≥10 hours had a higher risk of ADL disability than the reference group (7 hours per night). After adjusting for age, sex, and other potential confounders (Model 2), the association between 6 hours per night and ADL disability disappeared, while the associations between short (≤4 and 5 hours) and long (9 and ≥10 hours) sleep duration and ADL disability remained the same, with ORs of 1.91 (95% CI 1.60 - 2.27; *P*<.001), 1.33 (95% CI 1.09 - 1.62; *P*=.006), 1.48 (95% CI 1.13 - 1.93; *P*<.001), and 1.88 (95% CI 1.47 - 2.41; *P*<.001), respectively.

Table . Associations between sleep duration and activities of daily living (ADL) disability among participants from the 2018 China Health and Retirement Longitudinal Study.

	OR[a] (95% CI)	*P* value
Sleep duration per night, hours		
Model 1[b]		
≤4	2.85 (2.48 - 3.26)	<.001
5	1.51 (1.30 - 1.76)	<.001
6	1.17 (1.01 - 1.36)	.031
7	[Reference]	-[c]
8	1.08 (0.92 - 1.25)	.342
9	1.41 (1.14 - 1.75)	.002
≥10	2.03 (1.68 - 2.46)	<.001
Model 2[d]		
≤4	1.91 (1.60 - 2.27)	<.001
5	1.33 (1.09 - 1.62)	.006
6	1.11 (0.93 - 1.34)	.246
7	[Reference]	-[c]
8	1.06 (0.88 - 1.27)	.549
9	1.48 (1.13 - 1.93)	<.001
≥10	1.88 (1.47 - 2.41)	<.001

[a]OR: odds ratio.

[b]Model 1 was adjusted for age and sex.

[c]Not applicable.

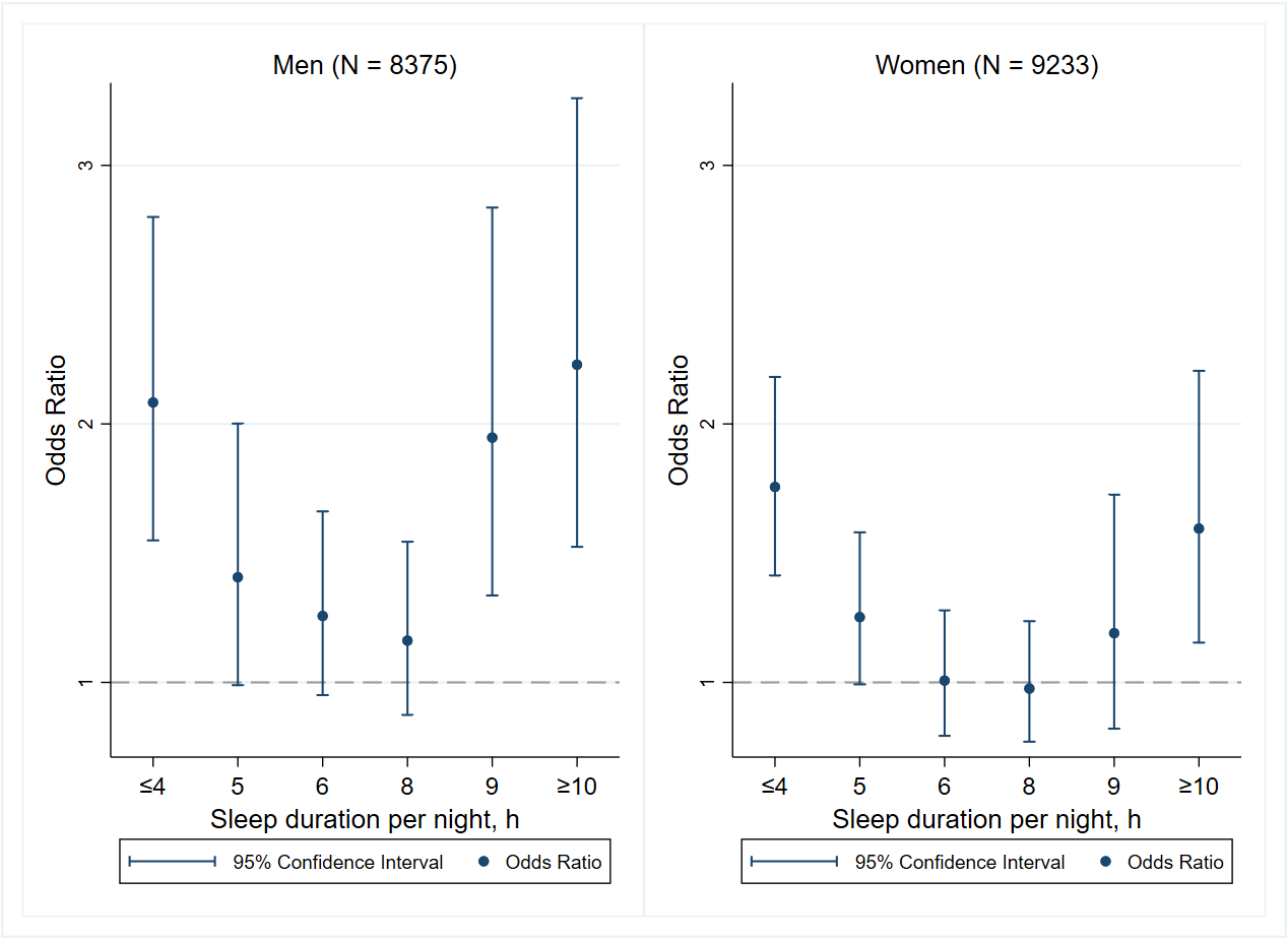
[d]Model 2 was adjusted for age, sex, education, marital status, tobacco use, alcohol use, afternoon napping, residence, depression, chronic disease conditions, visual impairment, hearing impairment, life satisfaction, self-reported health status, and sample weights.

Sensitivity Analyses

Figure 2 provides the results of stratified analysis by sex. Our findings found that for both men and women, short (≤4 hours) and long (9 and ≥10 hours) sleep duration were associated with higher ADL disability compared to the reference group. However, after adjusting for covariates, men with less than 4 hours (OR 2.08, 95% CI 1.55 - 2.80), 9 hours (OR 1.95, 95% CI13.4 - 2.84), and above 10 hours per night (OR 2.23, 95% CI 1.52 - 3.26) had a higher risk than women did.

XSL•FO
RenderX

Figure 2. Sex-specific effect of sleep duration on the activity of daily living disability.

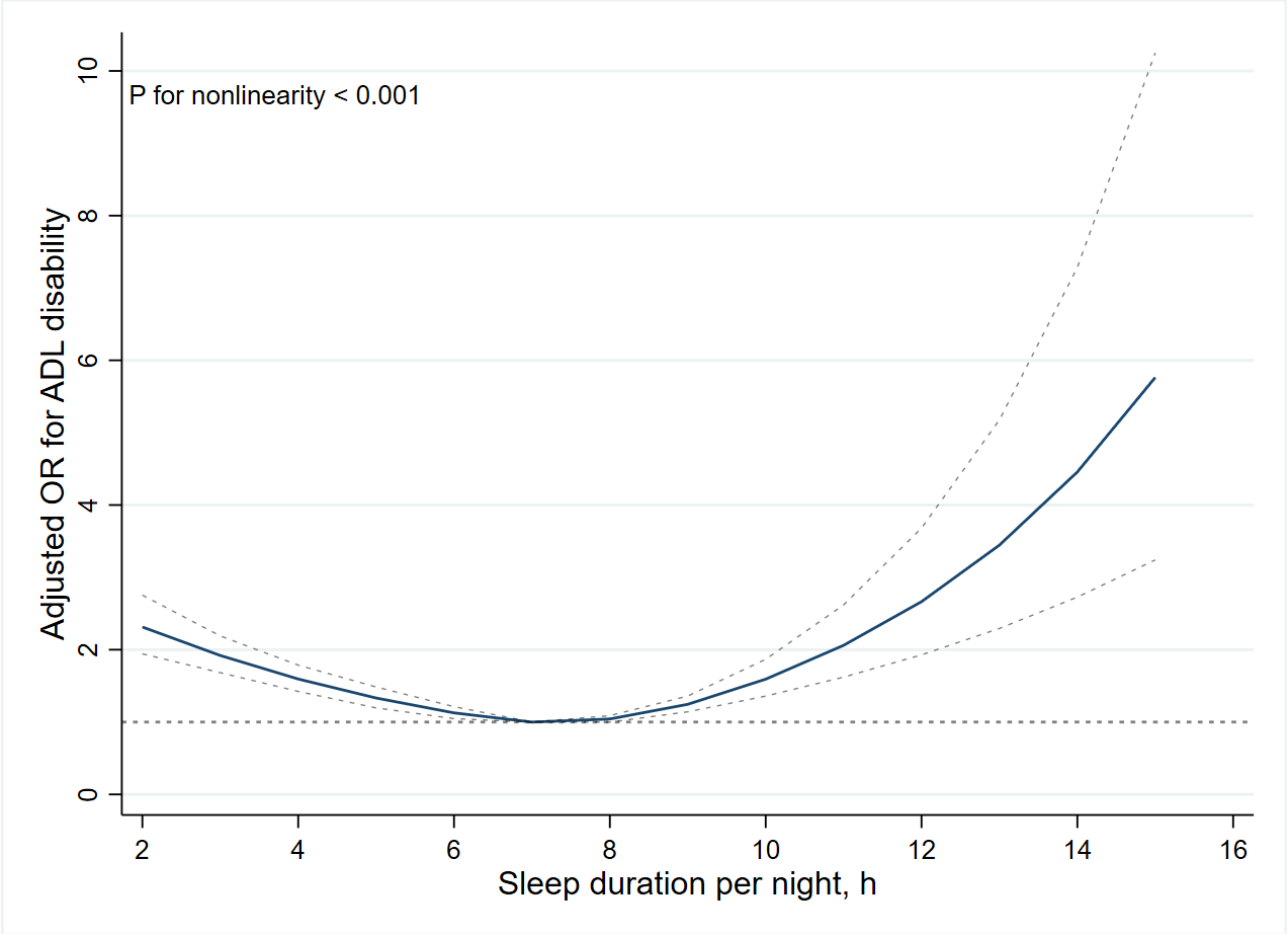


Nonlinear Relationship Between Sleep Duration and ADL Disability

In Figure 3, restricted cubic splines were used to flexibly model and visualize the relations between predicted sleep duration and ADL disability. Multivariable-adjusted restricted cubic splines analyses suggested U-shaped associations between sleep duration and ADL disability. The risk of ADL disability was negatively correlated with sleep duration until it bottomed out at 7 hours (OR 0.83, 95% CI 0.79 - 0.87; *P*<.001). However, when the sleep duration was higher than 7 hours, the risk of ADL disability increased significantly (OR 1.19, 95% CI 1.12 - 1.27; *P*<.001).

XSL•FO
RenderX

Figure 3. Nonlinear relationship of sleep duration and activity of daily living disability. OR: odds ratio; ADL: activities of daily living.



Discussion

Principal Findings and Comparison With Previous Works

In this nationally representative study, using the CHARLS database, a statistically significant U-shaped association was found between sleep duration and ADL disability. Individuals with short and long sleep duration were associated with a higher risk of ADL disability. To the best of our knowledge, the current study stands as one of the most extensive examinations of the correlation between sleep duration and ADL disability in the Chinese population.

Compared to other countries, the prevalence rates of ADL disability among older adults were higher in China [39], with one study observing a rate as high as 41% [40]. In contrast, in America, the percentage of older adults aged 65 - 74 years living with functional disability was 25.4% in 2015 [41]. The measurement of ADL disability serves as a crucial tool for monitoring the efficacy of therapeutic interventions across a range of diseases [42]. Given the well-established link between ADL disability and functional ability, it is important to conduct a study on the relationship between sleep duration and ADL disability.

Previous studies have estimated the correlation between sleep duration and ADL disability. The National Health & Aging Trends Study in the United States has demonstrated that insomnia symptoms were correlated with increased likelihood of functional limitations in later life [43]. The Centers for Disease Control and Prevention reported that adults who slept less than 7 hours per night were more prone to difficulties with daily activities than those who slept 7 - 9 hours per night [44]. Data from the New Integrated Suburban Seniority Investigation Project in Japan also suggested a heightened risk of incident disability in older adults associated with sleeping less than 6 hours per night [16]. However, this study found no significant link with long sleep duration, potentially due to its limited sample size (n=256) and subject detection bias. Conversely, data from 486,619 stroke survivors from the National Health Interview Survey revealed that long sleepers (≥9 hours) were more likely to experience problems with instrumental activities of daily living compared to average sleepers [13]. Our results are in accordance with the studies highlighting the increased risk of ADL disability associated with both short and long sleep. Meanwhile, our findings also align with a previous CHARLS study [45], which indicates that sleep durations of short (4 hours or less) and long (9 hours or more) predicted incident ADL disability.

Our results of multivariable-adjusted restricted cubic splines analyses also observed a U-shaped association between sleep duration and ADL disability. Our results lent support to the conclusion that adults aged 18 - 60 years were recommended to sleep at least 7 hours each night to foster optimal health and well-being, as suggested in previous studies [8,46]. Furthermore, “Healthy China 2030” echoes this advice, advocating for a sleep

XSL•FO
RenderX

duration of 7 - 8 hours per day for Chinese adults. Insufficient or excessive sleep duration per night was associated with an increased risk for coronary heart disease, cognitive decline, and functional limitations [30,45,47]. The evidence from China Kadoorie Biobank indicated that 23% of Chinese adults reported getting below 6 hours of sleep duration and 16% reported exceeding 10 hours [48]. The observed nonlinear association underscores the ongoing necessity for public education regarding sleep health and opportunities for healthcare providers to engage patients in discussions about the significance of sleep and to promote sleep health initiatives. Future research to prevent functional limitations may consider intervention strategies focused on sleep hygiene and inform the development of health policies.

The mechanisms by which sleep duration influences ADL disability remain unclear. Several plausible explanations have been proposed to explain the correlation between sleep duration and ADL disability. One possible explanation is tied to the pathology of age-related macular degeneration pathology. An exploratory study involving 277 patients with age-related macular degeneration found significant alterations in the expression levels of immediate early response 3 (IER-3), tissue inhibitor of metalloproteinase-3 (TIMP-3), beta 3-glucosyltransferase (B3GALTL), hepatic lipase (LIPC), and HtrA serine peptidase 1 (HTRA1) among individuals with sleep deprivation or patients with age-related macular degeneration who experienced an increase in sleep duration [49]. Meanwhile, physical activity markedly contributed to exacerbating the severity of age-related macular degeneration. Furthermore, IER-3 and TIMP-3 also had a role in regulating mechanisms in circadian rhythms [49]. The second explanation links the relationship between sleep duration and ADL disability to psychological distress [50]. Similarly, sleep duration also affected the relationship between ADL disability and psychological distress as a mediating pathway, which might also affect ADL disability by causing psychological distress [6]. However, the underlying mechanism governing the association between sleep duration and ADL disability remained uncertain.

In addition, our stratified analysis revealed an elevated risk of ADL disability among men with short and long sleep durations. Sex differences in sleep duration cannot be ignored, as men generally tend to sleep less than women throughout their lifespan [51]. Our findings align with those from the Swedish panel study of living conditions among the oldest old, which showed stronger associations between ADL and instrumental activities of daily living as well as other health indicators among men compared to women between 1992 and 2002 [52]. A longitudinal observation study of 3609 adults aged 65 - 89 years from the Korean National Health and Nutrition Examination Survey also found that men had 2 - 3 times higher odds of developing ADL disability than women [53]. One plausible explanation for this sex difference could be the more severe impact of conditions such as stroke, cancer, and pulmonary disease on older men compared to women, potentially due to men's greater propensity for alcohol and tobacco consumption, thereby increasing the greater risk of ADL disability.

Strengths and Limitations

Our study boasts significant strengths, notably its utilization of a nationally representative sample of Chinese adults aged 45 years and older, rather than relying on clinical samples or occupational cohorts. Furthermore, to our knowledge, it stands as one of the pioneering population-based studies to incorporate both sleep duration and ADL disability. However, several limitations of the current study merit consideration. First, due to its cross-sectional design, we are unable to establish causal relationships between sleep duration and ADL disability based on the observed associations, necessitating a cautious interpretation of the findings. Second, sleep duration was solely assessed through self-reporting, which may introduce potential biases. Third, the CHARLS questionnaire lacks comprehensive sleep-related information, such as objective assessments of the sleep quality, sleep disturbance, and difficulties initiating sleep, which could potentially act as a confounding factor, influencing our estimates. Finally, the binary classification of ADL disability, rather than a continuous scale, might underestimate the associations between sleep duration and ADL disability. Additionally, despite controlling for a range of covariates in the analysis, we cannot discount the influence of unknown factors in this study.

Conclusions

A nationally representative sample in China revealed a statistically significant U-shaped relationship between sleep duration and ADL disability. The U-shaped pattern underscores the importance of monitoring ADL disability in middle-aged and older individuals who exhibit either insufficient or excessive sleep duration. It is imperative for future longitudinal studies to elucidate the causal or bidirectional links between sleep duration and ADL disability, while also delving into the potential mechanisms underlying this intricate interplay.

Acknowledgments

This work was supported by Fundamental Research Funds for the Central Universities (No. 2022-JYB-PY-013). We thank the China Health and Retirement Longitudinal Study team for providing data and simple training in using the database.

Authors' Contributions

HR and XG provided the conception of the research design. HF conducted the data analysis and wrote the first version of the manuscript. WY, XG, and HR provided comments related to the presentation of the findings and critically reviewed the manuscript. All authors read and approved the final manuscript.

XSL•FO
RenderX

Conflicts of Interest

None declared.

References

1. Global report on health equity for persons with disabilities. World Health Organization. URL: https://www.who.int/teams/noncommunicable-diseases/sensory-functions-disability-and-rehabilitation/global-report-on-health-equity-for-persons-with-disabilities [accessed 2024-01-25]
2. He S, Craig BA, Xu H, et al. Unmet need for ADL assistance is associated with mortality among older adults with mild disability. J Gerontol A Biol Sci Med Sci 2015 Sep;70(9):1128-1132. [doi: 10.1093/gerona/glv028] [Medline: 25834196]
3. van Dalen-Kok AH, Pieper MJC, de Waal MWM, van der Steen JT, Scherder EJA, Achterberg WP. The impact of pain on the course of ADL functioning in patients with dementia. Age Ageing 2021 May 5;50(3):906-913. [doi: 10.1093/ageing/afaa247] [Medline: 33300044]
4. Falk Erhag H, Wetterberg H, Johansson L, Rydén L, Skoog I. Activities of daily living (ADL) and instrumental activities of daily living (IADL) disability in Swedish 85-year-olds born three decades apart-findings from the H70 study. Age Ageing 2021 Nov 10;50(6):2031-2037. [doi: 10.1093/ageing/afab112] [Medline: 34107021]
5. Štefan L, Vrgoč G, Rupčić T, Sporiš G, Sekulić D. Sleep duration and sleep quality are associated with physical activity in elderly people living in nursing homes. Int J Environ Res Public Health 2018 Nov 9;15(11):2512. [doi: 10.3390/ijerph15112512] [Medline: 30423981]
6. Xiao S, Shi L, Xue Y, et al. The relationship between activities of daily living and psychological distress among Chinese older adults: A serial multiple mediation model. J Affect Disord 2022 Mar 1;300:462-468. [doi: 10.1016/j.jad.2021.12.069] [Medline: 34954333]
7. Suzuki R, Meguro M, Meguro K. Sleep disturbance is associated with decreased daily activity and impaired nocturnal reduction of blood pressure in dementia patients. Arch Gerontol Geriatr 2011;53(3):323-327. [doi: 10.1016/j.archger.2010.11.006] [Medline: 21129792]
8. Liu Y, Wheaton AG, Chapman DP, Cunningham TJ, Lu H, Croft JB. Prevalence of healthy sleep duration among adults--United States, 2014. MMWR Morb Mortal Wkly Rep 2016 Feb 19;65(6):137-141. [doi: 10.15585/mmwr.mm6506a1] [Medline: 26890214]
9. Wang C, Bangdiwala SI, Rangarajan S, et al. Association of estimated sleep duration and naps with mortality and cardiovascular events: a study of 116 632 people from 21 countries. Eur Heart J 2019 May 21;40(20):1620-1629. [doi: 10.1093/eurheartj/ehy695] [Medline: 30517670]
10. Åkerstedt T, Ghilotti F, Grotta A, et al. Sleep duration and mortality - Does weekend sleep matter? J Sleep Res 2019 Feb;28(1):e12712. [doi: 10.1111/jsr.12712] [Medline: 29790200]
11. Zhou L, Yu K, Yang L, et al. Sleep duration, midday napping, and sleep quality and incident stroke: The Dongfeng-Tongji cohort. Neurology (ECronicon) 2020 Jan 28;94(4):e345-e356. [doi: 10.1212/WNL.0000000000008739] [Medline: 31827003]
12. World Mental Health Report: Transforming Mental Health for All: World Health Organization; 2022. URL: https://www.who.int/publications/i/item/9789240049338 [accessed 2025-02-21]
13. Seixas AA, Chung DP, Richards SL, et al. The impact of short and long sleep duration on instrumental activities of daily living among stroke survivors. Neuropsychiatr Dis Treat 2019;15:177-182. [doi: 10.2147/NDT.S177527] [Medline: 30655670]
14. Williams-Cooke C, Watts E, Bonnett J, Alshehri M, Siengsukon C. Association between sleep duration and functional disability in inpatient stroke rehabilitation: a pilot observational study. Arch Rehabil Res Clin Transl 2021 Sep;3(3):100150. [doi: 10.1016/j.arrct.2021.100150] [Medline: 34589700]
15. Yong LC, Li J, Calvert GM. Sleep-related problems in the US working population: prevalence and association with shiftwork status. Occup Environ Med 2017 Feb;74(2):93-104. [doi: 10.1136/oemed-2016-103638] [Medline: 27609325]
16. Zhang L, Ukawa S, Zhao W, et al. Daily sleep duration and the risk of incident disability among younger elderly Japanese adults in the New Integrated Suburban Seniority Investigation Project: A prospective study using competing event analysis. Geriatr Gerontol Int 2019 Sep;19(9):945-949. [doi: 10.1111/ggi.13732] [Medline: 31332921]
17. Campanini MZ, Mesas AE, Carnicero-Carreño JA, Rodríguez-Artalejo F, Lopez-Garcia E. Duration and quality of sleep and risk of physical function impairment and disability in older adults: results from the ENRICA and ELSA cohorts. Aging Dis 2019 Jun;10(3):557-569. [doi: 10.14336/AD.2018.0611] [Medline: 31165000]
18. Huang WC, Lin CY, Togo F, et al. Association between objectively measured sleep duration and physical function in community-dwelling older adults. J Clin Sleep Med 2021 Mar 1;17(3):515-520. [doi: 10.5664/jcsm.8964] [Medline: 33135630]
19. Sun R, Cao H, Zhu X, Liu JP, Dong E. Current aging research in China. Protein Cell 2015 May;6(5):314-321. [doi: 10.1007/s13238-015-0145-5] [Medline: 25779341]
20. Fang EF, Scheibye-Knudsen M, Jahn HJ, et al. A research agenda for aging in China in the 21st century. Ageing Res Rev 2015 Nov;24(Pt B):197-205. [doi: 10.1016/j.arr.2015.08.003] [Medline: 26304837]

XSL•FO
RenderX

21. Gu D, Sautter J, Pipkin R, Zeng Y. Sociodemographic and health correlates of sleep quality and duration among very old Chinese. Sleep 2010 May;33(5):601-610. [doi: 10.1093/sleep/33.5.601] [Medline: 20469802]
22. Chen X, Wang R, Zee P, et al. Racial/ethnic differences in sleep disturbances: the multi-ethnic study of atherosclerosis (MESA). Sleep 2015 Jun 1;38(6):877-888. [doi: 10.5665/sleep.4732] [Medline: 25409106]
23. Zhao Y, Hu Y, Smith JP, Strauss J, Yang G. Cohort profile: the China Health and Retirement Longitudinal Study (CHARLS). Int J Epidemiol 2014 Feb;43(1):61-68. [doi: 10.1093/ije/dys203] [Medline: 23243115]
24. Wu Z, Zhang H, Miao X, et al. High-intensity physical activity is not associated with better cognition in the elder: evidence from the China Health and Retirement Longitudinal Study. Alzheimers Res Ther 2021 Nov 3;13(1):182. [doi: 10.1186/s13195-021-00923-3] [Medline: 34732248]
25. 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. Int J Surg 2014 Dec;12(12):1495-1499. [doi: 10.1016/j.ijsu.2014.07.013] [Medline: 25046131]
26. Yao Y, Wang K, Xiang H. Association between cognitive function and ambient particulate matters in middle-aged and elderly Chinese adults: Evidence from the China Health and Retirement Longitudinal Study (CHARLS). Sci Total Environ 2022 Jul 1;828:154297. [doi: 10.1016/j.scitotenv.2022.154297] [Medline: 35288137]
27. Lin L, Wang HH, Lu C, Chen W, Guo VY. Adverse childhood experiences and subsequent chronic diseases among middle-aged or older adults in China and associations with demographic and socioeconomic characteristics. JAMA Netw Open 2021 Oct 1;4(10):e2130143. [doi: 10.1001/jamanetworkopen.2021.30143] [Medline: 34694390]
28. Bloomberg M, Brocklebank L, Hamer M, Steptoe A. Joint associations of physical activity and sleep duration with cognitive ageing: longitudinal analysis of an English cohort study. Lancet Healthy Longev 2023 Jul;4(7):e345-e353. [doi: 10.1016/S2666-7568(23)00083-1] [Medline: 37421962]
29. Jing R, Xu T, Rong H, Lai X, Fang H. Longitudinal association between sleep duration and depressive symptoms in Chinese elderly. Nat Sci Sleep 2020;12:737-747. [doi: 10.2147/NSS.S269992] [Medline: 33117009]
30. Ma Y, Liang L, Zheng F, Shi L, Zhong B, Xie W. Association between sleep duration and cognitive decline. JAMA Netw Open 2020 Sep 1;3(9):e2013573. [doi: 10.1001/jamanetworkopen.2020.13573] [Medline: 32955572]
31. Rong H, Wang X, Lai X, Yu W, Fei Y. Associations between sleep duration and sensory impairments among older adults in China. Front Aging Neurosci 2022;14:910231. [doi: 10.3389/fnagi.2022.910231] [Medline: 35754970]
32. Guo J, Fei Y, Li J, Zhang L, Luo Q, Chen G. Gender- and age-specific associations between sleep duration and prevalent hypertension in middle-aged and elderly Chinese: a cross-sectional study from CHARLS 2011-2012. BMJ Open 2016 Sep 6;6(9):e011770. [doi: 10.1136/bmjopen-2016-011770] [Medline: 27601494]
33. Song Y, Liu H, Gu K, Liu Y. Association between nighttime sleep duration, nap time, and mild cognitive impairment in Chinese older adults: a cross-sectional study. BMC Public Health 2024 Sep 2;24(1):2381. [doi: 10.1186/s12889-024-19900-0] [Medline: 39223492]
34. Wang Y, Jiang G, Hou N, et al. Effects and differences of sleep duration on the risk of new-onset chronic disease conditions in middle-aged and elderly populations. Eur J Intern Med 2023 Jan;107:73-80. [doi: 10.1016/j.ejim.2022.11.005]
35. Feng Z, Li Q, Zhou L, Chen Z, Yin W. The relationship between depressive symptoms and activity of daily living disability among the elderly: results from the China Health and Retirement Longitudinal Study (CHARLS). Public Health 2021 Sep;198:75-81. [doi: 10.1016/j.puhe.2021.06.023] [Medline: 34365109]
36. Chan S, Jia S, Chiu H, et al. Subjective health-related quality of life of Chinese older persons with depression in Shanghai and Hong Kong: relationship to clinical factors, level of functioning and social support. Int J Geriatr Psychiatry 2009 Apr;24(4):355-362. [doi: 10.1002/gps.2129] [Medline: 18773498]
37. Connolly D, Garvey J, McKee G. Factors associated with ADL/IADL disability in community dwelling older adults in the Irish longitudinal study on ageing (TILDA). Disabil Rehabil 2017 Apr;39(8):809-816. [doi: 10.3109/09638288.2016.1161848] [Medline: 27045728]
38. Hu Y, Zhou F, Kaminga AC, Yan S, Hu Z. Associations of depressive symptoms and chronic diseases with activities of daily living among middle-aged and older population in China: A population-based cohort study. Front Psychiatry 2022;13:848255. [doi: 10.3389/fpsyt.2022.848255]
39. Feng Q, Zhen Z, Gu D, Wu B, Duncan PW, Purser JL. Trends in ADL and IADL disability in community-dwelling older adults in Shanghai, China, 1998-2008. J Gerontol B Psychol Sci Soc Sci 2013 May;68(3):476-485. [doi: 10.1093/geronb/gbt012] [Medline: 23525547]
40. Zhang Y, Xiong Y, Yu Q, Shen S, Chen L, Lei X. The activity of daily living (ADL) subgroups and health impairment among Chinese elderly: a latent profile analysis. BMC Geriatr 2021 Jan 7;21(1):30. [doi: 10.1186/s12877-020-01986-x] [Medline: 33413125]
41. Wang T, Wu Y, Wang W, Zhang D. Association between coffee consumption and functional disability in older US adults. Br J Nutr 2021 Mar 28;125(6):695-702. [doi: 10.1017/S0007114520003153] [Medline: 32778181]
42. Lim YH, Baek Y, Kang SJ, Kang K, Lee HW. Clinical application of the experimental ADL test for patients with cognitive impairment: pilot study. Sci Rep 2021;11(1):356. [doi: 10.1038/s41598-020-78289-z]

XSL•FO
RenderX

43. Spira AP, Kaufmann CN, Kasper JD, et al. Association between insomnia symptoms and functional status in U.S. older adults. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69 Suppl 1(Suppl 1):S35-S41. [doi: 10.1093/geronb/gbu116] [Medline: 25342821]
44. Centers for Disease Control and Prevention (CDC). Effect of short sleep duration on daily activities--United States, 2005-2008. MMWR Morb Mortal Wkly Rep 2011 Mar 4;60(8):239-242. [Medline: 21368739]
45. Liu M, Du X, Sun Y, Zhou A, Sun S, Wu Y. The mediating role of cognition in the relationship between sleep duration and instrumental activities of daily living disability among middle-aged and older Chinese. Arch Gerontol Geriatr 2021;94:104369. [doi: 10.1016/j.archger.2021.104369] [Medline: 33556636]
46. Grandner MA, Chakravorty S, Perlis ML, Oliver L, Gurubhagavatula I. Habitual sleep duration associated with self-reported and objectively determined cardiometabolic risk factors. Sleep Med 2014 Jan;15(1):42-50. [doi: 10.1016/j.sleep.2013.09.012] [Medline: 24333222]
47. Lao XQ, Liu X, Deng HB, et al. Sleep quality, sleep duration, and the risk of coronary heart disease: a prospective cohort study With 60,586 adults. J Clin Sleep Med 2018 Jan 15;14(1):109-117. [doi: 10.5664/jcsm.6894] [Medline: 29198294]
48. Chen Y, Kartsonaki C, Clarke R, et al. Characteristics and correlates of sleep duration, daytime napping, snoring and insomnia symptoms among 0.5 million Chinese men and women. Sleep Med 2018 Apr;44:67-75. [doi: 10.1016/j.sleep.2017.11.1131]
49. Sharma K, Singh R, Sharma SK, Anand A. Sleeping pattern and activities of daily living modulate protein expression in AMD. PLoS ONE 2021;16(6):e0248523. [doi: 10.1371/journal.pone.0248523]
50. Zhang C, Xiao S, Lin H, et al. The association between sleep quality and psychological distress among older Chinese adults: a moderated mediation model. BMC Geriatr 2022;22(1):35. [doi: 10.1186/s12877-021-02711-y]
51. Jonasdottir SS, Minor K, Lehmann S. Gender differences in nighttime sleep patterns and variability across the adult lifespan: a global-scale wearables study. Sleep 2021 Feb 12;44(2):zsaa169. [doi: 10.1093/sleep/zsaa169] [Medline: 32886772]
52. Schön P, Parker MG, Kåreholt I, Thorslund M. Gender differences in associations between ADL and other health indicators in 1992 and 2002. Aging Clin Exp Res 2011 Apr;23(2):91-98. [doi: 10.1007/BF03351074] [Medline: 21743288]
53. Kim IH. Age and gender differences in the relation of chronic diseases to activity of daily living (ADL) disability for elderly South Koreans: based on representative data. J Prev Med Public Health 2011 Jan;44(1):32-40. [doi: 10.3961/jpmph.2011.44.1.32] [Medline: 21483221]

Abbreviations

ADL: activities of daily living
B3GALTL: beta 3-glucosyltransferase
CHARLS: China Health and Retirement Longitudinal Study
HTRA1: HtrA serine peptidase 1
IER-3: immediate early response 3
LIPC: hepatic lipase
OR: odds ratio
TIMP-3: tissue inhibitor of metalloproteinase-3
WHO: World Health Organization

Edited by Z Guo; submitted 04.08.24; peer-reviewed by E Pineda, SF Fung, S Ranjbaran; revised version received 19.01.25; accepted 19.01.25; published 05.03.25.

Please cite as:
Fan H, Yu W, Rong H, Geng X
Associations Between Sleep Duration and Activity of Daily Living Disability Among Older Adults in China: Cross-Sectional Study
Interact J Med Res 2025;14:e65075
URL: https://www.i-jmr.org/2025/1/e65075
doi:10.2196/65075

© Huimin Fan, Weijie Yu, Hongguo Rong, Xiaokun Geng. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 5.3.2025. 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

Multihealth Promotion Programs on Physical Health and Quality of Life in Older Adults: Quasi-Experimental Study

Li-Yun Lee[1], PhD, RN; Heng-Hsin Tung[2,3], PhD, RN, DNP, FNP; George Liao[4], BS; Su-Ju Liu[5], BS; Zi-Yu Chen[6], MSN; Yea-Ru Yang[7], PhD

[1]Department of Nursing, Dayeh University, Changhua, Taiwan

[2]Department of Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

[3]Tungs' Taichung MetroHarbor Hospital, Taichung, Taiwan

[4]YuSen Biotechnology Co, Ltd, Taichung City, Taiwan

[5]Keisei Pharmacy, Yunlin County, Taiwan

[6]YuSen Biotechnology Co, Ltd, Taichung, Taiwan

[7]Department of Physical Therapy and Assistive Technology, National Yang Ming Chiao Tung University, Taipei, Taiwan

Corresponding Author:
Heng-Hsin Tung, PhD, RN, DNP, FNP
Department of Nursing
National Yang Ming Chiao Tung University
No 155, Sec 2, Linong St
Beitou Dist
Taipei, 112
Taiwan
Phone: 886 2 2826 7000 ext 67991
Fax: 886 228250472
Email: shannontung719@gmail.com

Abstract

Background: Physical activity and appropriate nutrition are essential for older adults. Improving physical health and quality of life can lead to healthy aging.

Objective: This study aims to investigate the long-term effects of multihealth promotion programs on the physical and mental health of older adults in communities.

Methods: A quasi-experimental method was used to recruit 112 older adults voluntarily from a pharmacy in central Taiwan between April 2021 and February 2023. Participants were divided into an experimental group receiving a multihealth promotion program and a control group with no specific intervention. The study measured frailty, nutritional status, well-being, and quality of life using standardized tools such as the Clinical Frailty Scale (CFS), Mini-Nutritional Assessment-Short Form (MNA-SF), Well-being Scale for Elders, and the EQ-5D-3L. Data were analyzed using descriptive statistics, independent *t* tests, Pearson correlation, and generalized estimating equations.

Results: A total of 112 participants were recruited. There were 64 (57.1%) in the experimental group and 48 (42.9%) in the control group. The experimental group exhibited significantly better quality of life (EQ-5D index) at weeks 12 (β=–.59; *P*=.01) and 24 (β=–.44; *P*=.04) compared to the control group. The experimental group muscle mass significantly increased at weeks 24 (β=4.29; *P*<.01) and 36 (β=3.03; *P*=.01). Upper limb strength improved significantly at weeks 12 (β=3.4; *P*=.04) and 36 (β=5; *P*=.01), while core strength showed significant gains at weeks 12 (β=4.43; *P*=.01) and 36 (β=6.99; *P*<.01). Lower limb strength increased significantly only at week 12 (β=4.15; *P*=.01). Overall physical performance improved significantly at weeks 12 (β=5.47; *P*<.01), 24 (β=5.17; *P*<.01), and 36 (β=8.79; *P*<.01).

Conclusions: The study's findings highlight the practical benefits of interventions, including physical and social activities and nutritional support, in enhancing the quality of life and general physical health of older adults. This study's findings have significant implications for clinical practice. These findings can aid in the establishment of effective interventions for older adults.

Trial Registration: ClinicalTrials.gov NCT05412251; https://clinicaltrials.gov/study/NCT05412251

(Interact J Med Res 2025;14:e65213) doi:10.2196/65213

XSL•FO
RenderX

KEYWORDS

older adult; body composition; physical activity; health promotion; exercise; nutrition; diet; well-being; quality-of-life; QoL; gerontology; geriatrics

Introduction

As the challenges of global aging intensify, it is projected that by 2050, individuals aged 60 years and older will account for 22% of the global population [1]. This demographic shift brings a surge in health concerns among older adults, necessitating increased attention. Aging is commonly associated with significant changes in body composition, including substantial loss of muscle strength and mass, which elevates the risk of sarcopenia and functional decline [2]. In addition, older adults are particularly vulnerable to moderate-to-severe malnutrition, further exacerbating health deterioration and reducing quality of life (QoL) [3]. Consequently, health promotion strategies should prioritize improving the nutritional status of older adults and encouraging regular physical activity to support overall health and QoL [4]. Research consistently demonstrates that older adults who engage in regular physical activity experience significantly better physical and mental health outcomes than sedentary individuals, underscoring the undeniable benefits of exercise interventions [5]. As such, exercise interventions are widely regarded as a practical approach to enhancing QoL in older adults [6].

Exercise interventions for older adults encompass a variety of methods, including yoga, breathing exercises, meditation, Baduanjin, Western gymnastics, 24-form Tai Chi, aerobic exercises, outdoor activities, balance, strength, and resistance training [4,5,7]. These interventions typically last 12 weeks, with sessions ranging from 30 to 90 minutes, 2 to 5 times per week, with 3 sessions per week being the most common frequency. However, systematic reviews and meta-analyses reveal that current health promotion studies predominantly focus on single exercise modalities, yielding inconsistent effects on physical function and lacking in-depth exploration of multicomponent exercise interventions, particularly in practical applications and comprehensive outcome assessments [4,5,7]. Specifically, Frost et al [7] conducted a meta-analysis of 7 randomized controlled trials (RCTs), of which only one combined exercise with nutritional interventions. This study found that strength training supplemented with protein significantly improved muscle mass. However, the effects on psychological health and well-being were not adequately explored, highlighting the need for further research on the holistic benefits of multicomponent interventions. Similarly, Iwano et al [8] emphasized in their systematic review that the variation in intervention types (eg, art therapy and social activities) and measurement tools (eg, well-being scales and depression indices) in studies on older adults' well-being limits the comparability and generalizability of intervention effects, underscoring the necessity for further research. Wei et al [5] advocated for future studies incorporating diverse forms of physical activity to investigate their effects on physical and mental health systematically. They also encouraged older adults to engage in various activities to enhance QoL effectively.

In addition, Sirikul et al [4] conducted a meta-analysis of multicomponent exercise interventions (combining strength, balance, flexibility, and endurance training) and nutritional interventions among frail older adults aged 65 years and above in community settings. Their findings suggest that combining nutritional interventions, such as protein supplementation, with exercise interventions, particularly strength training, effectively improves older adults' muscle mass and physical function. These findings underscore the potential value of multicomponent strategies and emphasize the importance of integrating exercise and nutritional interventions in future health promotion programs to enhance older adults' overall health and functional outcomes.

Although exercise interventions are widely recognized for their potential to improve QoL in older adults, their effects on physical function and psychological health remain inconsistent. For example, Sirikul et al [4] conducted a meta-analysis of 21 exercise intervention studies, reporting predominantly positive effects on muscle strength and balance. However, their findings also highlighted the lack of comprehensive evaluations of multicomponent interventions (eg, combining exercise and nutrition). Frost et al [7] reported that only one of 7 RCTs included combined exercise and nutritional interventions, showing that strength training with protein supplementation significantly increased muscle mass. Nevertheless, the effects of this combined intervention on mental and psychological health were not thoroughly explored. Marcos-Pardo PJ et al [9] found that moderate-to-high intensity resistance circuit training significantly improved muscle strength in older adults but had no significant impact on QoL. Conversely, Chittrakul et al [10] reported that multisystem exercise training significantly enhanced QoL in older adults. Existing studies suggest that exercise interventions should be tailored to the physical condition of older adults. Combining social participation, autonomy enhancement, nutritional interventions, physical training, and cognitive activities in multicomponent health promotion programs improves nutritional status and overall physical function. It effectively enhances QoL and well-being while reducing fall risks [4,5,7].

Well-being, a subjective experience characterized by happiness or satisfaction, is a core goal universally pursued by humans. It is closely associated with reduced risks of depression, diminished feelings of hopelessness, improved mental health, enhanced QoL, and strengthened family relationships. For older adults, well-being often stems from life achievements and successful experiences [11]. Enhancing well-being helps older adults maintain health and encourages them to adopt a more positive outlook on life. However, previous research has predominantly focused on treating and managing physical diseases in older adults, with relatively little attention given to well-being [12]. Furthermore, significant disparities remain in the findings of studies investigating the effects of interventions on older adults' well-being [8]. Studies have shown that social activities significantly improve psychological health by engaging older adults in meaningful, hopeful, and enjoyable interactions

XSL•FO
RenderX

[13,14]. Maintaining QoL in older adults is essential for individual health and reflects their adaptability and resilience during aging. High QoL effectively reduces mental health issues, promotes active community participation, strengthens social connections, and alleviates feelings of loneliness [6].

In addition, smart devices provide precise data and enhance older adults' engagement and adherence to exercise [15,16]. However, further research is needed to compare the effects of specific multicomponent intervention models. This study investigates the comprehensive effects of a multihealth promotion program combining smart exercise devices, social activities, and nutritional prescriptions on the physical function, QoL, and well-being of older adults in community settings. The long-term benefits will also be tracked to provide evidence-based recommendations for promoting healthy aging.

Methods

Design and Sample

This study adopted a quasi-experimental design, with participants voluntarily recruited from a pharmacy located in central Taiwan between April 2021 and February 2023. The participants themselves determined allocation to the experimental and control groups. The trial design was registered with ClinicalTrials.gov (identifier: NCT05412251) and followed the CONSORT (Consolidated Standards of Reporting Trials) guidelines (Multimedia Appendix 1). The inclusion criteria for this study were participants (1) aged 65 years or older and (2) fluent in Mandarin (speaking or reading). The study excluded participants who had a history of cognitive impairment, cancer, myocardial infarction, or congestive heart failure. This study used G power software to estimate the sample size. The effect size was calculated based on the pre- and posttest scores of the quality of life of older adults in the intervention group and the control group in a previous study [17]. The effect size was 0.19, the detection power was 0.8, the significance level (α) was .05, and the number of measurements was four times. The results showed that 40 participants were required in each group to ensure that a statistically significant difference in quality of life between the two groups could be detected during the follow-up period. Considering the attrition rate in interventional studies can be as high as 40% [18], at least 56 participants were recruited in each group at baseline. Data were collected at baseline (T0) and 12 weeks, 24 weeks, and 36 weeks postintervention (T1, T2, and T3). The data was collected face-to-face and self-reported by participants.

Multihealth Promotion Program Intervention

The design of the multihealth promotion program was based on previous studies. We combined 3 intervention strategies: social activities, physical training, and nutritional support, and used artificial intelligence exercise equipment as an assessment tool to measure physical performance during the exercise process [5,13,19]. For 12 weeks, participants in the experimental group engaged in social and physical activities while receiving nutritional support. (1) Social activities: this was combined with activities held at community base C, such as makeup, reading picture books, art classes, board games, and more, for at least 120 minutes, once a week. (2) Physical training: the utilization of artificial intelligence exercise equipment, suitable for middle-aged and older people's exercise training and approved by the Food and Drug Administration of the Ministry of Health and Welfare. This was carried out at least 60 minutes, 3 times a week, including 8-10 minutes of warm-up exercise before exercise, after exercise, aerobic exercise, muscle strengthening training, and balance training, each requiring 15-20 minutes. Before each exercise, researchers used the exercise assessment and screening for you (EASY) tool to assess participants' exercise risks, and the AI-enabled exercise equipment will evaluate participants and provide appropriate exercise prescriptions. Throughout the exercise, the researchers observed the physical condition of the participants constantly. If there were signs of discomfort, participants were asked to stop immediately. (3) Nutritional support: nutritionists screened participants' nutritional status based on the mini nutrition assessment tool and based on personal wishes, provided free protein nutrition products in this plan. During the entire study period, the control group did not receive the intervention program for social activities, physical activity, or nutritional support. They only took measurements and questionnaires at monitoring time points (baseline, 12, 24, and 36 weeks). The on-site nutritionist was available for any questions from the control group.

Measurements

The study measured demographic and baseline characteristics, frailty, nutrition status, well-being, and quality of life.

Demographic Data Form

A researcher-made demographic form was prepared based on a literature review. It included age, gender, BMI, education level, lived conditions, and chronic disease.

Clinical Frailty Scale

The Clinical Frailty Scale (CFS) is used to assess frailty in older adults and was developed by Rockwood et al in 2005. This scale was updated to version 2.0 in 2020. The frailty scale ranges from 1 to 9, where level 1 indicates very fit and level 9 indicates terminally ill, corresponding to a survival time of fewer than 6 months [20].

Mini-Nutritional Assessment-Short Form

The Mini-Nutritional Assessment-Short Form (MNA-SF) consists of 6 questions that assess key areas of nutritional assessment: decline in food intake, weight loss, mobility, psychological stress or acute disease, neuropsychological problems, and BMI. The total score classifies individuals into 3 categories, normal nutritional status (12-14 points), at-risk of malnutrition (8-11 points), and malnourished (0-7 points). The MNA-SF has been validated in various studies and settings, demonstrating high sensitivity and specificity [21].

The EASY Tool

The EASY is an exercise screening and assessment tool designed to help older adults choose safe and effective exercise modalities and physical activity programs. There are 6 questions, each of which can be answered with "yes" or "no." If the answer to one of the questions is "yes," the participant must find a health care

XSL•FO
RenderX

provider to discuss an appropriate exercise plan before exercising [22].

Well-Being Scale for Elders

The Well-being Scale for Elders is a 9-item self-report questionnaire, including emotional, psychological, and social well-being. The higher scores indicate better well-being. It has good reliability and construct validity and can be effectively used to measure the happiness of older adults. The Cronbach α is 0.91, and the factor analysis results explain 59.67% of the variation [12]. In this study, Cronbach α is 0.82.

EQ-5D-3L Instrument

The EQ-5D 3L is a standardized instrument developed by the EuroQol Group to measure general health status. It is widely used in both clinical and economic evaluations to assess health-related quality of life. The EQ-5D index comprises 5 dimensions: mobility, self-care, daily activities, pain or discomfort, and anxiety or depression. Each dimension has 3 levels of severity (no problems, moderate problems, and extreme problems). Higher scores on this index indicate a poorer health-related quality of life. EQ-VAS (European Quality of Life Visual Analogue Scale) is a visual analog scale ranging from 0 to 100, with higher scores indicating a better-perceived quality of life. Respondents evaluate their overall health on a scale from 0, indicating the worst possible health, to 100, indicating the best possible health [23]. The EQ-5D-3L has demonstrated good reliability, validity, and sensitivity for measuring health-related quality of life [24]. In this study, Cronbach α is 0.8.

Statistical Analysis

All data were analyzed using the SPSS software program (version 24; IBM Corp). This research uses descriptive statistics to analyze frequencies, means, and SD. Differences in outcome variables at baseline between the two groups are analyzed using the chi-square test and an independent *t* test. The Generalized Estimating Equations (GEE) analyze changes over time, group differences, and group-by-time interactions, including body composition, frailty, well-being, and quality of life.

Ethical Considerations

This study was conducted with the utmost ethical considerations and approved by the National University Institutional Review Board (YM110168EF). All participants were respected, and their autonomy was upheld, as they provided written informed consent before enrollment. Participants were fully informed by a research assistant about the study's purpose, procedure, benefits, and potential risks, ensuring transparency and respect for their rights and welfare. The participant data has been deidentified for privacy concerns.

Results

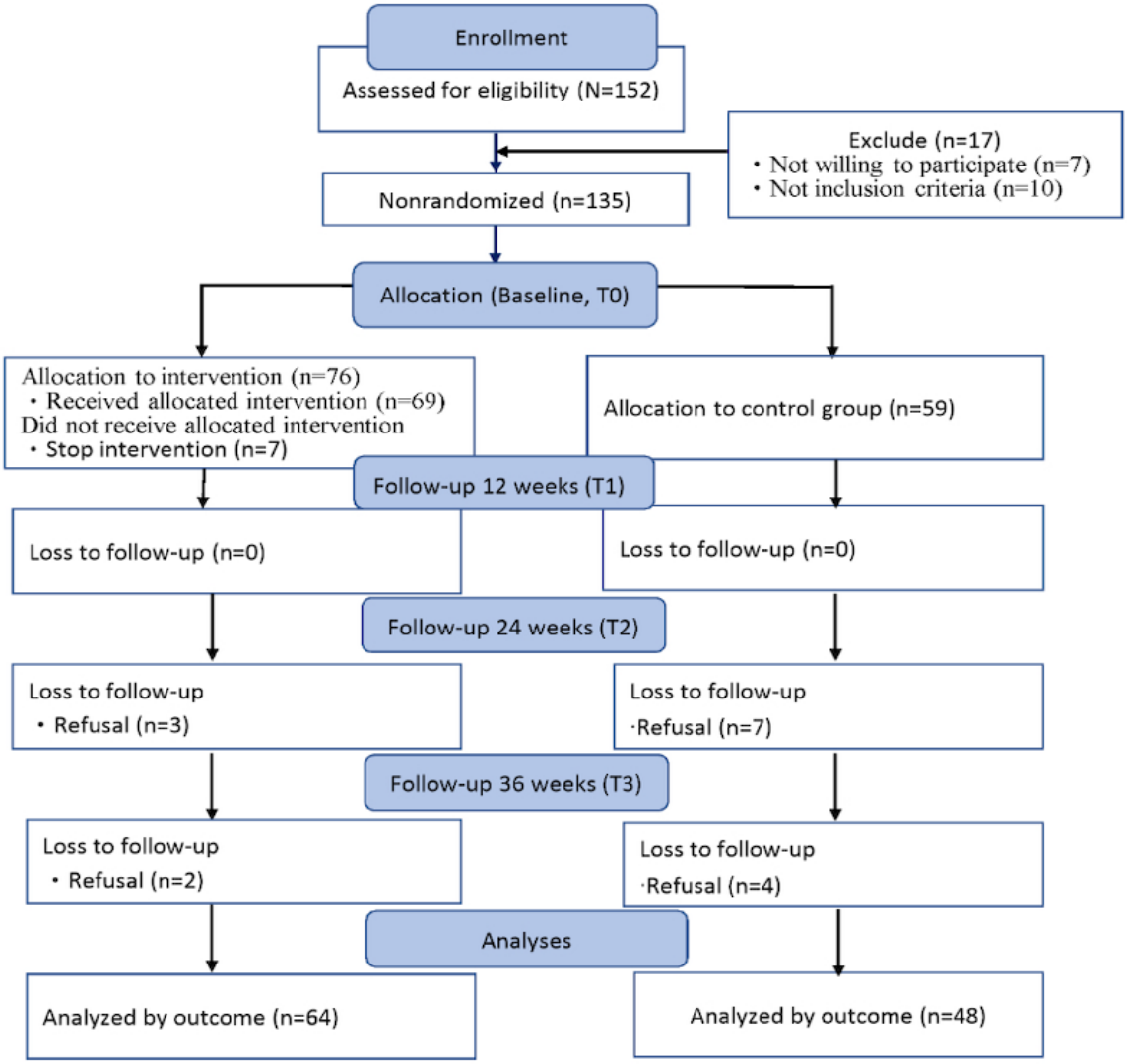
In this study, after excluding ineligible participants and those who dropped out, a total of 112 older adults (64 in the experimental group and 48 in the control group) completed the study, with 15.8% and 18.6% attrition rates, respectively (Figure 1). Most participants were female (67%), with an average BMI of 24.5 (3.6) kg/m². Overall, 23.2% (26/112) of the participants lived alone; the average frailty score was 2.2 (SD 0.9). Baseline characteristics showed good homogeneity between the 2 groups. As of February 2023, we have recruited 152 participants (Table 1).

Table 2 reveals average baseline BMI in the control group as 24.13 (SD 3.89) kg/m², which slightly decreased to 24.06 (SD 3.63) kg/m² at week 36. Frailty scores rose from 2.31 to 2.45, while the average well-being scores improved from 38.21 (SD 5.42) to 38.59 (SD 5.45). Quality-of-life scores showed improvement, decreasing from an average of 5.50 (SD 1.07) to 5.35 (SD 0.90). In the experimental group, average baseline BMI was 24.79 (SD 3.34) kg/m², which reduced to 24.43 (SD 3.42) kg/m² by week 36. Frailty scores increased slightly from 2.11 to 2.16, and average well-being scores rose from 37.86 (SD 5.99) to 38.91 (SD 4.97). Quality-of-life scores improved significantly, decreasing from an average of 5.98 (SD 1.46) to 5.42 (SD 0.79).

Table 3 shows that the GEE analysis revealed significant BMI reductions across all participants in the first 12 weeks. At week 36, the experimental group had a significantly lower BMI compared to the control group. Notably, BMI reductions were more pronounced in the experimental group at weeks 12 (β=–.22; *P*=.02) and 24 (β=–.20; *P*=.06), with a smaller decrease at week 36 (β=.15; *P*=.23). The interaction effect between groups at week 36 was significant (β=–.51; *P*<.001), highlighting a greater BMI reduction in the experimental group. Frailty assessment showed no significant interaction between time and group, suggesting stable frailty levels during the intervention. Similarly, no significant differences in well-being were observed. However, the experimental group exhibited significantly better quality of life (EQ-5D index) at weeks 12 (β=–.59; *P*=.01) and 24 (β=–.44; *P*=.04) compared to the control group. No significant differences were observed in EQ-VAS scores between groups.

XSL•FO
RenderX

Figure 1. Flowchart of participants' recruitment.



XSL•FO
RenderX

Table 1. Baseline characteristics (N=112).

Variables	Total (N=112)	EG[a] (n=64)	CG[b] (n=48)	Values (χ^2/*t*)	*P* value
Age (year), mean (SD)[a]	75.1 (6.2)	75.81 (5.1)	74.16 (1.1)	–1.32	.19[c]
Sex, n (%)				4.36	.04[d]
Female	75 (67)	48 (75)	27 (56.3)		
Male	37 (33)	16 (25)	21 (43.8)		
BMI (kg/m^2), mean (SD)	24.5 (3.6)	24.8 (3.3)	24.1 (3.9)	–0.96	.34[c]
Education level, n (%)				4.82	.19[d]
None	13 (11.6)	7 (10.9)	6 (12.5)		
Elementary	56 (50)	36 (56.3)	20 (41.7)		
Junior high	10 (8.9)	7 (10.9)	3 (6.3)		
Senior high and above	33 (29.5)	14 (21.9)	19 (39.6)		
Lived conditions, n (%)				1.19	.28[d]
Alone	26 (23.2)	13 (20.3)	6 (12.5)		
With family	86 (76.8)	51 (79.7)	42 (87.5)		
Chronic disease, n (%)				3.04	.08[d]
No	86 (76.8)	53 (82.8)	33 (68.8)		
Yes	26 (23.2)	11 (17.2)	15 (31.3)		
Frailty[e], mean (SD)	2.2 (0.9)	2.1 (0.9)	2.3 (0.9)	1.18	.24[c]
Well-being[f], mean (SD)	38 (5.7)	37.7 (6)	38.2 (5.4)	0.32	.75[c]
EQ-5D index, mean (SD)	5.7 (1.3)	5.9 (1.4)	5.5 (1.1)	–1.72	.09[c]
EQ-VAS[g], mean (SD)	80.6 (9.7)	81.2 (8.2)	79.7 (11.3)	–0.84	.4[c]

[a]EG: experimental group.

[b]CG: control group.

[c]Conducted using independent *t* test (*df*=110).

[d]Conducted using χ^2 test (*df*=1,3).

[e]Frailty: The Clinical Frailty Scale.

[f]Well-being: Well-being scale for Elders.

[g]EQ-VAS: European Quality of Life Visual Analogue Scale.

XSL•FO
RenderX

Table 2. Changes at different periods between the experimental group and the control group (N=112).

Variable and group	Baseline, mean (SD)	12 Weeks, mean (SD)	24 Weeks, mean (SD)	36 Weeks, mean (SD)
BMI (kg/m^2)				
EG[a]	24.79 (3.34)	24.49 (3.46)	24.39 (3.40)	24.43 (3.42)
CG[b]	24.13 (3.89)	23.91 (3.77)	23.93 (3.73)	24.06 (3.63)
Frailty[c]				
EG	2.11 (0.88)	2.06 (0.81)	2.12 (0.85)	2.16 (0.84)
CG	2.31 (0.93)	2.46 (0.80)	2.46 (1.01)	2.45 (1.02)
Well-being[d]				
EG	37.86 (5.99)	37.53 (4.92)	38.92 (4.92)	38.91 (4.97)
CG	38.21 (5.42)	37.65 (5.70)	38.33 (4.82)	38.59 (5.45)
EQ–5D index				
EG	5.98 (1.46)	5.50 (0.81)	5.42 (0.85)	5.42 (0.79)
CG	5.50 (1.07)	5.60 (1.14)	5.38 (0.96)	5.35 (0.90)
EQ-VAS[e]				
EG	81.23 (8.24)	80.63 (8.30)	82.42 (10.11)	83.20 (9.16)
CG	79.69 (11.30)	80.35 (10.68)	81.33 (10.14)	81.74 (10.04)

[a]EG: experimental group.

[b]CG: control group.

[c]Frailty: the Clinical Frailty Scale.

[d]Well-being: Well-being scale for Elders.

[e]EQ-VAS: European Quality of Life Visual Analogue Scale.

XSL•FO
RenderX

Table 3. Generalized estimating equations analysis of BMI, frailty, happiness, and quality of life in older participants (N=112).

Parameter	β	SE	95% Wald CI	*P* value
BMI (kg/m^2)				
(Intercept)	24.13	0.56	23.04 to 25.22	<.001
EG[a]	0.66	0.69	–0.69 to 2.02	.34
CG[b] (reference group)	0[a]	—[c]	—	—
36 weeks	0.15	0.12	–0.10 to 0.39	.23
24 weeks	–0.20	0.10	–0.40 to 0.01	.06
12 weeks	–0.22	0.09	–0.41 to –0.03	.02
Baseline	0[a]	—	—	—
EG*36 weeks	–0.51	0.15	–0.81 to –0.21	.001
EG*24 weeks	–0.21	0.14	–0.48 to 0.07	.14
EG*12 weeks	–0.08	0.13	–0.33 to 0.17	.53
EG* baseline	0[a]	—	—	—
Frailty[d]				
(Intercept)	2.31	0.13	2.05 to 2.57	.001
EG	–0.20	0.17	–0.54 to 0.13	.24
CG (reference group)	0[a]	—	—	—
36 weeks	0.15	0.09	–0.04 to 0.33	.13
24 weeks	0.15	0.09	–0.03 to 0.32	.1
12 weeks	0.15	0.10	–0.05 to 0.34	.14
Baseline	0[a]	—	—	—
EG*36 weeks	–0.10	0.12	–0.34 to 0.15	.43
EG*24 weeks	–0.07	0.11	–0.29 to 0.15	.54
EG*12 weeks	–0.19	0.12	–0.43 to 0.04	.11
EG* baseline	0[a]	—	—	—
Well-being[e]				
(Intercept)	38.21	0.77	36.69 to 39.73	<.001
EG	–0.35	1.07	–2.45 to 1.76	.75
CG (reference group)	0[a]	—	—	—
36 weeks	0.30	0.81	–1.28 to 1.88	.71
24 weeks	0.13	0.87	–1.58 to 1.83	.89
12 weeks	–0.56	0.82	–2.18 to 1.05	.49
Baseline	0[a]	—	—	—
EG*36 weeks	0.75	1.12	–1.44 to 2.94	.50
EG*24 weeks	0.94	1.20	–1.41 to 3.28	.43
EG*12 weeks	0.23	1.03	–1.79 to 2.26	.82
EG* baseline	0[a]	—	—	—
EQ-5D index				
(Intercept)	5.55	0.15	5.2 to 5.8	.001
EG	0.48	0.24	0.02 to 0.95	.04

XSL•FO
RenderX

Parameter	β	SE	95% Wald CI	*P* value
CG (reference group)	0[a]	—	—	—
36 weeks	–0.16	0.16	–0.47 to 0.14	.30
24 weeks	–0.13	0.15	–0.42 to 0.17	.41
12 weeks	0.10	0.11	–0.11 to 0.32	.35
Baseline	0[a]	—	—	—
EG*36 weeks	–0.40	0.23	–0.84 to 0.04	.08
EG*24 weeks	–0.44	0.21	–0.85 to –0.03	.04
EG*12 weeks	–0.59	0.19	–0.96 to –0.21	.01
EG* baseline	0[a]	—	—	—
EQ-VAS[f]				
(Intercept)	79.69	1.61	76.52 to 82.85	.01
EG	1.55	1.91	–2.19 to 5.29	.42
CG (reference group)	0[a]	—	—	—
36 weeks	1.74	1.46	–1.11 to 4.6	.23
24 weeks	1.64	1.04	–0.39 to 3.68	.11
12 weeks	0.66	0.71	–0.73 to 2.06	.35
Baseline	0[a]	—	—	—
EG*36 weeks	0.22	1.96	–3.62 to 4.07	.91
EG*24 weeks	–0.46	1.81	–4 to 3.08	.80
EG*12 weeks	–1.28	1.23	–3.68 to 1.13	.30
EG* baseline	0[a]	—	—	—

[a]EG: experimental group.

[b]CG: control group.

[c]Not applicable.

[d]Frailty: the Clinical Frailty Scale.

[e]Well-being: Well-being scale for Elders.

[f]EQ-VAS: European Quality of Life Visual Analogue Scale.

Table 4 shows that GEE analysis revealed no significant changes in body fat percentage at weeks 12 (β=–.03; *P*=.69) and 36 (β=–.31; *P*=.27). However, muscle mass significantly increased at weeks 24 (β=4.29; *P*<.001) and 36 (β=3.03; *P*=.01), demonstrating the intervention's effectiveness in enhancing muscle mass. Upper limb strength improved significantly at weeks 12 (β=3.40; *P*=.04) and 36 (β=5; *P*=.01), while core strength showed significant gains at weeks 12 (β=4.43; *P*=.01) and 36 (β=6.99; *P*<.001). Lower limb strength increased significantly only at week 12 (β=4.15; *P*=.01). Overall physical performance improved significantly at weeks 12 (β=5.47; *P*<.001), 24 (β=5.17; *P*<.001), and 36 (β=8.79; *P*<.001).

XSL•FO
RenderX

Table 4. Body composition in the experimental group analyzed by generalized estimating equations (n=64).

Outcome	B	SE	95% CI	Wald χ^2	*P*
Body fat percentage					
Intercept	3.77	0.27	3.24 to 4.29	196.56	<.001
36 weeks[a]	–0.31	0.28	–0.86 to 0.24	1.24	.27
24 weeks[a]	–0.42	0.28	–0.97 to 0.13	2.27	.13
12 weeks[a]	–0.03	0.08	–0.19 to 0.13	0.15	.69
Muscle content					
Intercept	33.06	1.2	30.7 to 35.42	753.74	<.001
36 weeks[a]	3.03	1.24	0.61 to 5.47	6	.01
24 weeks[a]	4.29	1.21	1.92 to 6.66	12.58	<.001
12 weeks[a]	–0.67	0.73	–2.10 to 0.77	0.83	.36
Upper limb score					
Intercept	65.91	1.36	63.25 to 68.57	2363.36	<.001
36 weeks[a]	5	1.98	1.13 to 8.86	6.43	.01
24 weeks[a]	2.42	1.93	–1.37 to 6.20	1.57	.21
12 weeks[a]	3.4	1.67	0.12 to 6.68	4.13	.04
Core score					
Intercept	65.05	1.1	62.9 to 67.2	3520.3	<.001
36 weeks[a]	6.99	2	3.07 to 10.91	12.2	<.001
24 weeks[a]	1.02	1.80	–2.5 to 4.54	0.33	.57
12 weeks[a]	4.43	1.44	1.6 to 7.25	9.45	.01
Lower limb score					
Intercept	67.48	1.35	64.84 to 70.12	2508.61	<.001
36 weeks[a]	3.23	2	–0.7 to 7.15	2.6	.11
24 weeks[a]	0.91	1.96	–2.94 to 4.76	0.22	.64
12 weeks[a]	4.15	1.4	1.42 to 6.88	8.85	.01
Comprehensive sports score					
Intercept	62.38	1.05	60.32 to 64.43	3553.19	<.001
36 weeks[a]	8.79	1.36	6.12 to 11.46	41.58	<.001
24 weeks[a]	5.17	1.21	2.79 to 7.55	18.1	<.001
12 weeks[a]	5.47	1.37	2.78 to 8.16	15.83	<.001

[a]Reference group is the baseline.

Discussion

Principal Findings

This study's findings indicate that an intervention program encompassing exercise, social activities, and nutritional prescriptions significantly improved muscle mass, muscle strength, and QoL in the experimental group. Exercise programs have been demonstrated to enhance physical function, body composition, QoL, and muscle strength in older adults [6,9,10,16]. However, their effects on mental health and subjective QoL are less consistent and warrant further investigation. In this study, the experimental group's QoL (measured by the EQ-5D index) was significantly higher than the control group at weeks 12 and 24. However, no significant difference was observed in EQ-VAS scores. By week 36, the experimental group showed a trend of improvement in QoL, suggesting that the intervention may have long-term benefits for mental health. These findings align with Marcos-Pardo et al [6,9], who reported that resistance training interventions

XSL•FO
RenderX

improved lean body mass, muscle strength, and QoL. In a randomized controlled trial (RCT) by P J Marcos-Pardo et al [6], the experimental group participated in an 8-week outdoor resistance training program involving bi-weekly sessions with weight machines. Results revealed significant increases in lean body mass, reductions in fat mass, and improvements in muscle strength (arm and leg), functional activity, and QoL. At the same time, the control group experienced declines in muscle strength, activity capacity, dynamic balance, and QoL. Another RCT by Marcos-Pardo et al [9] demonstrated that older adults in the experimental group engaging in moderate-to-high intensity resistance circuit training for 12 weeks experienced significant reductions in body weight, fat mass, and BMI, along with significant improvements in lean body mass, functional independence, and upper and lower limb strength. However, no significant changes in body fat or visceral fat were observed in the experimental group during the study period. Chittrakul et al [10] conducted an RCT involving prefrail older adults with a 12-week exercise intervention and a two-year follow-up. The experimental group exhibited significant improvements in muscle strength at week 12, with enhancements in knee extension, balance, and reaction time persisting through week 24. The experimental group's health-related QoL was significantly higher than the control group at week 12. However, by week 24, despite the control group engaging in thrice-weekly stretching exercises, there was no significant difference in QoL between the groups, contrasting with this study's sustained improvements in the experimental group beyond week 24. This study also observed a BMI reduction at week 36, differing from previous findings [9], underscoring the need to explore further BMI changes and their relationships with muscle mass and strength.

Unlike the QoL findings, this study did not detect significant differences in well-being between the experimental and control groups, contrasting with the findings of Lai et al [25] and Jung et al [26]. The quasi-experimental study by Lai et al [25] demonstrated that a 6-day outdoor activity program, including kayaking, rafting, hiking, and camping, significantly improved well-being, self-image, and mental health among older adults, differing from this study's results. Similarly, Jung et al [26] reported that a 6-week integrated cognitive function improvement program significantly enhanced cognitive function and well-being in the experimental group, with no significant changes observed in the control group. The lack of significant well-being improvements in this study might be attributable to participant characteristics, cultural contexts, or the predominantly indoor nature of the intervention activities. Notably, previous research [27] has highlighted the mediating and moderating roles of outdoor and group exercise in enhancing well-being, emphasizing the psychological benefits of these activities. This underscores the importance of environmental and social factors in promoting well-being. The findings related to well-being suggest that additional intervention components beyond physical and nutritional strategies may be necessary. Factors such as depression risk, social relationships, and leisure activities have been shown to influence well-being [11]. Furthermore, Yang et al [27] emphasized that outdoor environments and group interactions significantly enhance well-being by fostering social connections and providing meaningful engagement. Future studies should consider integrating these elements to maximize the psychological benefits of multicomponent interventions.

Exercise and nutritional interventions have positively affected older adults' physical function and weight management. Liao et al [19] showed that combining exercise with nutritional support improved grip strength and walking speed despite limited changes in lean body mass. Similarly, in this study, while there were no significant changes in body fat or visceral fat in the experimental group, muscle mass significantly increased at week 24, with effects sustained through week 36. Upper limb strength improved significantly up to week 36, whereas lower limb strength showed significant improvements only at week 12. These findings highlight the need for more targeted designs in future interventions to address lower limb training.

This study underscores the potential of multidisciplinary active aging programs to improve physical function, QoL, and muscle strength among community-dwelling older adults. Such programs have been shown to reduce frailty risk, enhance flexibility, and promote social participation [28]. However, single-component nutritional interventions may have limited effects on individuals without malnutrition or unintentional weight loss [4], consistent with this study's findings. Clinicians should prioritize integrating physical, nutritional, and social interventions into comprehensive strategies tailored to older adults' needs. Muscle mass and muscle strength are positively correlated, and increasing muscle mass is key to improving muscle strength in older adults. Muscle mass and body fat percentage are important predictors of muscle strength. With aging, physical fitness declines, and body composition changes (eg, decreased muscle mass and increased obesity) are common [29]. Enhancing muscle strength is critical for clinical health care providers, contributing to improved stability, coordination, flexibility, and balance, ultimately reducing injury risk [30]. As global life expectancy rises and the proportion of older adults increases, promoting healthy aging is not merely a slogan but an urgent priority. However, physical inactivity and nutritional issues are prevalent among older adults [31], and these challenges are closely tied to their health and QoL [3]. While this study did not delve deeply into the influencing factors of QoL, well-being, and nutritional status, the results suggest these areas are important directions for future research.

Limitations

This study has several limitations that must be noted. First, the quasi-experimental design may introduce selection bias, as participants were not randomized. Second, the predominantly indoor nature of the activities may have limited the potential psychological benefits. Third, this study did not explore the influence of individual variability in adherence and baseline activity levels on the outcomes. Fourth, it was unable to determine the potential for individuals in either the experimental or control group to participate in additional outdoor or physical activities outside of the intervention program, which could have influenced the results. Further studies should address these limitations by employing randomized controlled trial designs, which include outdoor and group-based activities, examining

XSL•FO
RenderX

the effects of participant adherence and baseline characteristics, and tracking any additional activities performed outside of the intervention.

Conclusion

In conclusion, this study confirms the beneficial effects of multicomponent intervention programs on older individuals' physical function and quality of life. However, the limited benefits on well-being underline the importance of more comprehensive solutions that consider psychological and social aspects. Future studies should concentrate on the long-term assessment of nutritional status, quality of life, and well-being and the creation of individualized intervention programs that incorporate outdoor activities to promote active and healthy aging.

Acknowledgments

The authors thank YuSen Biotechnology Co., Ltd. and Keisei Pharmacy, which led to the successful completion of this research project. This study was funded by YuSen Biotechnology Co., Ltd. and Keisei Pharmacy.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

All authors contributed to the conceptualization, writing—review and editing, and approval of the submitted manuscript and will be accountable for its contents. HHT and GL supervised the study. HHT and LYL curated the data. HHT, GL, YRY, SJL, and ZYC performed the investigation and methodology. LYL conducted the formal analysis. HHT, GL, and LYL handled project administration. GL and SJL managed resources. LYL and HHT performed visualization. LYL and HHT wrote the original draft. GL acquired the funding.

Conflicts of Interest

The authors declare that this study was financially supported by YuSen Biotechnology Co., Ltd. and Keisei Pharmacy. Among the authors 2 coauthors are employees of YuSen Biotechnology Co., Ltd. While efforts were made to maintain research integrity and objectivity, this relationship may be considered a potential conflict of interest.

Multimedia Appendix 1
CONSORT 2010 checklist.
[PDF File (Adobe PDF File), 220 KB - ijmr_v14i1e65213_app1.pdf]

References

1. Ageing and health. World Health Organization. 2022. URL: https://www.who.int/news-room/fact-sheets/detail/ageing-and-health [accessed 2022-10-01]
2. Rocha JNDS, Pereira-Monteiro MR, Vasconcelos ABS, Pantoja-Cardoso A, Aragão-Santos JC, Da Silva-Grigoletto ME. Different resistance training volumes on strength, functional fitness, and body composition of older people: a systematic review with meta-analysis. Arch Gerontol Geriatr 2024;119:105303. [doi: 10.1016/j.archger.2023.105303] [Medline: 38128241]
3. Katsas K, Mamalaki E, Kontogianni MD, Anastasiou CA, Kosmidis MH, Varlamis I, et al. Malnutrition in older adults: correlations with social, diet-related, and neuropsychological factors. Nutrition 2020;71:110640. [doi: 10.1016/j.nut.2019.110640] [Medline: 31812000]
4. Sirikul W, Buawangpong N, Pinyopornpanish K, Siviroj P. Impact of multicomponent exercise and nutritional supplement interventions for improving physical frailty in community-dwelling older adults: a systematic review and meta-analysis. BMC Geriatr 2024;24(1):958 [FREE Full text] [doi: 10.1186/s12877-024-05551-8] [Medline: 39558234]
5. Wei L, Hu Y, Tao Y, Hu R, Zhang L. The effects of physical exercise on the quality of life of healthy older adults in China: A systematic review. Front Psychol 2022;13:895373 [FREE Full text] [doi: 10.3389/fpsyg.2022.895373] [Medline: 35800916]
6. Marcos-Pardo PJ, Espeso-García A, Vaquero-Cristóbal R, Abelleira-Lamela T, González-Gálvez N. The effect of resistance training with outdoor fitness equipment on the body composition, physical fitness, and physical health of middle-aged and older adults: a randomized controlled trial. Healthcare (Basel) 2024;12(7):726 [FREE Full text] [doi: 10.3390/healthcare12070726] [Medline: 38610148]
7. Frost R, Belk C, Jovicic A, Ricciardi F, Kharicha K, Gardner B, et al. Health promotion interventions for community-dwelling older people with mild or pre-frailty: a systematic review and meta-analysis. BMC Geriatr 2017;17(1):157 [FREE Full text] [doi: 10.1186/s12877-017-0547-8] [Medline: 28728570]
8. Iwano S, Kambara K, Aoki S. Psychological interventions for well-being in healthy older adults: systematic review and meta-analysis. J Happiness Stud 2022;23(5):2389-2403. [doi: 10.1007/s10902-022-00497-3]

XSL•FO
RenderX

9. Marcos-Pardo PJ, Orquin-Castrillón FJ, Gea-García GM, Menayo-Antúnez R, González-Gálvez N, Vale RGDS, et al. Effects of a moderate-to-high intensity resistance circuit training on fat mass, functional capacity, muscular strength, and quality of life in elderly: a randomized controlled trial. Sci Rep 2019;9(1):7830 [FREE Full text] [doi: 10.1038/s41598-019-44329-6] [Medline: 31127163]
10. Chittrakul J, Siviroj P, Sungkarat S, Sapbamrer R. Multi-system physical exercise intervention for fall prevention and quality of life in pre-frail older adults: a randomized controlled trial. Int J Environ Res Public Health 2020;17(9):3102 [FREE Full text] [doi: 10.3390/ijerph17093102] [Medline: 32365613]
11. Segura A, Cardona D, Segura A, Robledo CA, Muñoz DI. The subjective perception of the happiness of older adult residents in Colombia. Front Med (Lausanne) 2023;10:1055572 [FREE Full text] [doi: 10.3389/fmed.2023.1055572] [Medline: 37215723]
12. Xiao Y, Song D, Fu N, Zhang L, Zhang Y, Shen R, et al. Effects of resistance training on sarcopenia in patients with intestinal failure: a randomized controlled trial. Clin Nutr 2023;42(10):1901-1909. [doi: 10.1016/j.clnu.2023.07.013]
13. Park JH, Kang SW. Social interaction and life satisfaction among older adults by age group. Healthcare (Basel) 2023;11(22):2951 [FREE Full text] [doi: 10.3390/healthcare11222951] [Medline: 37998443]
14. Trudel-Fitzgerald C, Millstein RA, von Hippel C, Howe CJ, Tomasso LP, Wagner GR, et al. Psychological well-being as part of the public health debate? Insight into dimensions, interventions, and policy. BMC Public Health 2019;19(1):1712 [FREE Full text] [doi: 10.1186/s12889-019-8029-x] [Medline: 31856772]
15. Shang M, Dedeyne L, Dupont J, Vercauteren L, Amini N, Lapauw L, et al. Otago exercises monitoring for older adults by a single IMU and hierarchical machine learning models. IEEE Trans Neural Syst Rehabil Eng 2024;32:462-471. [doi: 10.1109/TNSRE.2024.3355299] [Medline: 38231806]
16. Broekhuizen K, de Gelder J, Wijsman CA, Wijsman LW, Westendorp RGJ, Verhagen E, et al. An internet-based physical activity intervention to improve quality of life of inactive older adults: a randomized controlled trial. J Med Internet Res 2016;18(4):e74 [FREE Full text] [doi: 10.2196/jmir.4335] [Medline: 27122359]
17. Serra-Prat M, Sist X, Domenich R, Jurado L, Saiz A, Roces A, et al. Effectiveness of an intervention to prevent frailty in pre-frail community-dwelling older people consulting in primary care: a randomised controlled trial. Age Ageing 2017;46(3):401-407. [doi: 10.1093/ageing/afw242] [Medline: 28064172]
18. Ga J, Tj W. Systematic review and meta analysis of differential attrition between active and control arms in randomized controlled trials of lifestyle interventions in chronic disease. BMC Med Res Methodol 2021;21(1):122 [FREE Full text] [doi: 10.1186/s12874-021-01313-x] [Medline: 34126934]
19. Liao X, Cheng D, Li J, Zhu L, Zhang S, Jing X, et al. Effects of oral oligopeptide preparation and exercise intervention in older people with sarcopenia: a randomized controlled trial. BMC Geriatr 2024;24(1):260 [FREE Full text] [doi: 10.1186/s12877-024-04860-2] [Medline: 38500043]
20. Rockwood K, Theou O. Using the clinical frailty scale in allocating scarce health care resources. Can Geriatr J 2020;23(3):210-215 [FREE Full text] [doi: 10.5770/cgj.23.463] [Medline: 32904824]
21. Kaiser MJ, Bauer JM, Ramsch C, Uter W, Guigoz Y, Cederholm T, MNA-International Group. Validation of the mini nutritional assessment short-form (MNA-SF): a practical tool for identification of nutritional status. J Nutr Health Aging 2009;13(9):782-788 [FREE Full text] [doi: 10.1007/s12603-009-0214-7] [Medline: 19812868]
22. Resnick B, Ory MG, Hora K, Rogers ME, Page P, Bolin JN, et al. A proposal for a new screening paradigm and tool called exercise assessment and screening for you (EASY). J Aging Phys Act 2008;16(2):215-233. [doi: 10.1123/japa.16.2.215] [Medline: 18483443]
23. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
24. Sun LF, Huang S, Li YF, Yang ZQ, Yang XJ, Zou JY, et al. Health-related quality of life measured by EQ-5D-3L for the spouses of breast cancer patients. Front Oncol 2022;12 [FREE Full text] [doi: 10.3389/fonc.2022.983704] [Medline: 36330472]
25. Lai DWL, Ou X, Jin J. A quasi-experimental study on the effect of an outdoor physical activity program on the well-being of older Chinese people in Hong Kong. Int J Environ Res Public Health 2022;19(15):8950 [FREE Full text] [doi: 10.3390/ijerph19158950] [Medline: 35897322]
26. Jung E, Choi Y, Lee K. Effects of integrative cognitive function improvement program on cognitive function, oral health, and mental health in older people: A randomized clinical trial. Int J Environ Res Public Health 2022;19(21) [FREE Full text] [doi: 10.3390/ijerph192114339] [Medline: 36361215]
27. Yang Q, Tang Y, Jennings G, Zhao B, Zhu F, Ma X. Physical activity and subjective well-being of older adults during COVID-19 prevention and control normalization: Mediating role of outdoor exercise environment and regulating role of exercise form. Front Psychol 2022;13:1014967 [FREE Full text] [doi: 10.3389/fpsyg.2022.1014967] [Medline: 36337512]
28. Chi YC, Wu CL, Liu HT. Effect of a multi-disciplinary active aging intervention among community elders. Medicine (Baltimore) 2021;100(51):e28314 [FREE Full text] [doi: 10.1097/MD.0000000000028314] [Medline: 34941126]
29. Hiol AN, von Hurst PR, Conlon CA, Mugridge O, Beck KL. Body composition associations with muscle strength in older adults living in Auckland, New Zealand. PLoS One 2021;16(5):e0250439 [FREE Full text] [doi: 10.1371/journal.pone.0250439] [Medline: 34048458]

XSL•FO
RenderX

30. Mak HW, Noguchi T, Bone JK, Wels J, Gao Q, Kondo K, et al. Hobby engagement and mental wellbeing among people aged 65 years and older in 16 countries. Nat Med 2023;29(9):2233-2240 [FREE Full text] [doi: 10.1038/s41591-023-02506-1] [Medline: 37696932]
31. Tcymbal A, Abu-Omar K, Hartung V, Bußkamp A, Comito C, Rossmann C, et al. Interventions simultaneously promoting social participation and physical activity in community living older adults: a systematic review. Front Public Health 2022;10:1048496 [FREE Full text] [doi: 10.3389/fpubh.2022.1048496] [Medline: 36568739]

Abbreviations

CFS: Clinical Frailty Scale
CONSORT: Consolidated Standards of Reporting Trials
EASY: exercise assessment and screening for you
EQ-VAS: European Quality of Life Visual Analogue Scale
GEE: Generalized Estimating Equations
MNA-SF: Mini-Nutritional Assessment-Short Form
QoL: quality of life
RCT: randomized clinical trial

Edited by T de Azevedo Cardoso; submitted 12.08.24; peer-reviewed by M Gasmi; comments to author 26.11.24; revised version received 31.01.25; accepted 10.03.25; published 01.05.25.

Please cite as:
Lee LY, Tung HH, Liao G, Liu SJ, Chen ZY, Yang YR
Multihealth Promotion Programs on Physical Health and Quality of Life in Older Adults: Quasi-Experimental Study
Interact J Med Res 2025;14:e65213
URL: https://www.i-jmr.org/2025/1/e65213
doi:10.2196/65213
PMID:40310677

©Li-Yun Lee, Heng-Hsin Tung, George Liao, Su-Ju Liu, Zi-Yu Chen, Yea-Ru Yang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 01.05.2025. 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

Predicting Age-Related Hearing Loss in Community-Dwelling Older Adults: Multicenter Retrospective Cohort Study

Jing Li[1], MS; Shuai Jin[1], PhD; Liu Sun[1], PhD; Jun-E Liu[1], PhD; Qiang Shen[2], MS; Miao Shang[3], MS; Hanting Wang[1], BS; Yuanyuan Zhao[1], BS

[1]School of Nursing, Capital Medical University, Beijing, Beijing, China

[2]Dongcheng District Community Health Service Management Center, Beijing, China

[3]Department of Otolaryngology, Beijing Tongren Hospital, Beijing, China

Corresponding Author:
Jun-E Liu, PhD
School of Nursing
Capital Medical University
No. 10, Xitoutiao, You'anmenwai
Fengtai District
Beijing, Beijing, 100069
China
Phone: 86 185 0013 6128
Email: liujune66@ccmu.edu.cn

Abstract

Background: Age-related hearing loss (ARHL) is associated with severe negative outcomes, including social isolation, depression, and cognitive decline. Despite this, routine ARHL screening is often neglected in primary care due to low awareness, resource limitations, and inefficiencies. A practical risk assessment tool could effectively address this gap.

Objective: This study aims to develop and validate a user-friendly nomogram for identifying older adults at high risk of ARHL in community settings, thereby facilitating targeted screening and timely interventions.

Methods: This multicenter retrospective cohort study included 34,983 older adults from 3 primary health care centers in Beijing (January 2020 to October 2023). Data from center A (n=18,707) were used for model development, with external validation performed on cohorts from center B (n=11,008) and center C (n=5268). Least absolute shrinkage and selection operator and logistic regression identified the final predictors. Model performance was evaluated using discrimination, calibration, and decision curve analysis, leading to the development of an online nomogram.

Results: In the training cohort (center A), 1177 participants (6.3%) had hearing loss. Six key predictors were identified: age, education, exercise frequency, physical function, dietary habits, and hypertension. The multivariate logistic regression model demonstrated good discrimination in internal validation (area under the curve [AUC] 0.806, 95% CI 0.782-0.831; sensitivity 0.774; specificity 0.820). External validation confirmed its generalizability (AUC 0.720, 95% CI 0.670-0.771 and AUC 0.747, 95% CI 0.712-0.782). Decision curve analysis highlighted a substantial clinical net benefit. A user-friendly online prediction web page was also developed.

Conclusions: We successfully developed and validated a dynamic, online nomogram for predicting ARHL in older adults. Comprising 6 readily available predictors, this model shows potential as a practical, online tool for proactive risk identification in primary care. However, further validation in larger and more diverse populations is essential to confirm its generalizability and real-world clinical utility.

(Interact J Med Res 2025;14:e81135) doi:10.2196/81135

KEYWORDS

primary health care; age-related hearing loss; prediction model; multicenter study; nomogram

XSL•FO
RenderX

Introduction

Age-related hearing loss (ARHL) is the third most prevalent health condition among older adults, after hypertension and arthritis, and has become a significant public health issue [1]. The World Health Organization estimates that more than 1.5 billion people worldwide experience hearing loss, a number projected to reach 2.5 billion by 2050 [2]. In China, hearing loss affects 30% of adults older than 60 years and rises to 50% for those older than 70 years, totaling at least 70 million cases [3]. ARHL can hinder communication and social interaction, leading to greater risks of social isolation, depression, cognitive decline [4-6], frailty, and falls [7,8].

Despite these negative health consequences, routine hearing screening remains significantly underused, with rates below 25% in older populations [9,10]. Key barriers include limited patient awareness and financial constraints; for example, a diagnostic audiogram in China can cost ¥100 (US $14.1) to ¥300 (US $42.3), a substantial out-of-pocket expense for many older adults; these challenges are further compounded by limited time, training, and equipment among primary care clinicians [11-14]. Furthermore, high false-positive rates from some hearing screening methods can increase costs, prompting policymakers to adopt a cautious approach [3,15].

Leveraging primary health care centers for the management of hearing loss is the most suitable and cost-effective approach [16,17]. A preliminary prediction tool designed for these centers could accurately identify individuals with high risk for targeted audiometric screening, thereby optimizing referrals and specialist resources. While previous models have been developed using known risk factors for hearing impairment, such as age, environmental noise [18], comorbidities [19], exercise, and nutrition [20], their clinical applicability is often limited by small sample sizes, a lack of external validation, and reliance on complex variables.

Therefore, this study aimed to develop a streamlined nomogram-based prediction model for ARHL among community-dwelling older adults, using a large sample derived from health check-ups conducted in primary health care institutions in Beijing, China. Additionally, the model was externally validated in 2 independent datasets. An online tool was used to deploy the model as an interactive application, enabling primary health care providers to perform risk prediction without requiring any software installation or programming expertise. This study adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis guidelines [21].

Methods

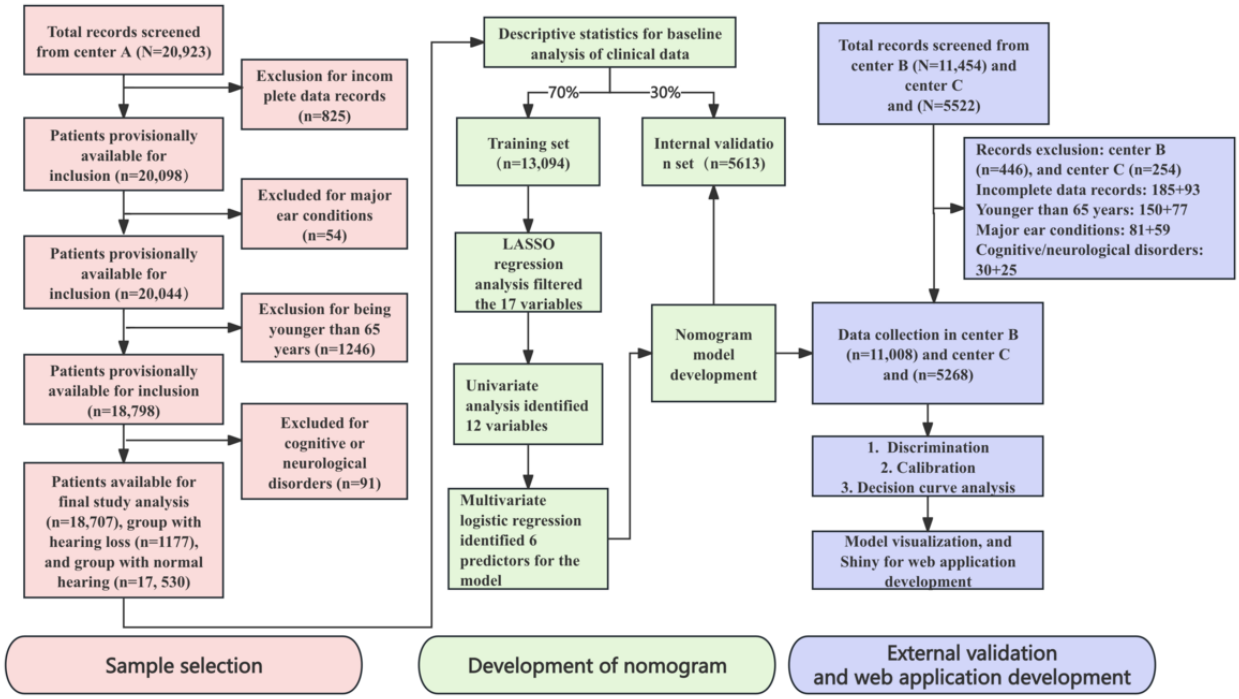
Study Population

This multicenter retrospective cohort study used electronic health examination records from 3 community primary health care centers (center A, center B, and center C) in Beijing, China. Data were collected from older adults who participated in the annual health checkup under the National Public Health Service Project between January 1, 2020, and October 30, 2023. The dataset included comprehensive information on demographics, medical history, physical examination results, and chronic disease records. For this study, the dataset from center A, which had the largest sample size, was used for model development and internal validation, while data from center B and center C were used for external validation. A detailed study design and sample selection are shown in Figure 1.

From center A, a total of 20,923 patient records were initially screened. First, 3.94% (825/20,923) records with incomplete data were excluded. After subsequently excluding 0.26% (54/20,923) individuals with major ear conditions, 5.96% (1246/20,923) individuals aged younger than 65 years, and 0.43% (91/20,923) individuals with cognitive or neurological disorders, a final cohort of 89.41% (18,707/20,923) participants was included in the analysis. The external validation cohorts were derived from center B and center C. Initially, 11,454 records from center B and 5522 records from center C were screened. After applying the exclusion criteria, a total of 3.89% (446/11,454) participants were excluded from center B and 4.60% (254/5522) from center C. This process resulted in final validation cohorts of 96.11% (11,008/11,454) participants from center B and 95.40% (5268/5522) from center C.

XSL•FO
RenderX

Figure 1. Flowchart for hearing loss prediction model development and validation.



Assessment of Hearing Loss

The outcome variable for this study was hearing loss, which was assessed using the whispered voice test as specified by the standardized protocol in the National Basic Public Health Service Standards (Third Edition) [22], issued by China's National Health Commission. This screening was performed by trained primary health care professionals to ensure consistency across all centers.

Specifically, the whispered voice test, a widely recognized and validated screening tool, was used due to its simplicity, cost-effectiveness, and reliability in detecting hearing loss in clinical and community settings [23,24]. The protocol dictates the following procedure: the examiner stands approximately 0.5 m from the participant's tested ear, positioned slightly behind them to prevent lip-reading. The participant's nontested ear is occluded (blocked) by the examiner to isolate the tested ear. After a full exhalation, the examiner whispers a simple, standardized phrase (eg, What is your name?). Failure to correctly repeat the phrase was classified as the presence of hearing loss. For the purpose of our model, this binary outcome (presence vs absence of hearing loss) was used as the dependent variable [22].

Sample Size Estimation

The determination of an adequate sample size for a multivariate predictive model requires a sufficient representation of individuals and outcome events relative to predictor parameters [25]. The combined sample size from 3 primary health care institutions provided robust support for using 33 predictive variables in both modeling and validation. Specifically, according to the widely applied 10 events per variable principle, at least 330 events (33 variables×10 events) are required [26]. Given an event rate of 0.063, this translates to approximately 5239 cases. The modeling dataset comprised 18,707 cases and 1177 events, which is more than 3 times the minimum requirement of 330 events. This large number of events per variable provides sufficient statistical power and minimizes the risk of model overfitting.

Model Development and Validation

Variable selection was conducted using the least absolute shrinkage and selection operator (LASSO) regression, which applies a Least Absolute Deviations penalty to shrink some regression coefficients to 0, simplifying the model while preserving variables with strong predictive power. Tenfold cross-validation determined the optimal penalty parameter λ (λ_min), and variables with nonzero coefficients in the final model were chosen as candidate predictors [27,28]. Next, univariate analysis was performed on the LASSO-selected variables. Variables with P values <0.05 were considered for inclusion in the multivariate logistic regression model. A forward stepwise regression approach was applied to identify the final predictors to be included in the model. To handle missing data, a complete case analysis was used. This approach, which excludes records with any missing data on key variables, was considered appropriate because the proportion of such records was low across all cohorts (3.94% in center A, 1.62% in center B, and 1.68% in center C), consistently remaining below 5%.

Statistical analysis was performed using R software (version 4.2.2; R Foundation for Statistical Computing). To facilitate clinical application of the prediction model, we developed an interactive online application using the *Shiny* framework (RStudio, PBC). The final logistic regression model and coefficients were embedded into the server logic using the *rms* package to ensure consistency with the nomogram. A user-friendly interface was created with *Shiny*, *shiny Widgets*,

XSL•FO
RenderX

and *shiny themes* packages, allowing real-time input and calculation of predicted risk. Dynamic visualizations of individualized risk were implemented using *plotly*, providing immediate graphical feedback. The application was deployed on a cloud-based *Shiny* server to allow online access without requiring local installation.

Model validation included 3 aspects: discrimination, calibration, and clinical utility [29]. Discrimination was assessed using the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and Matthews correlation coefficient to evaluate the model's ability to distinguish between individuals with hearing loss and those without hearing loss. The calibration of the model was assessed using calibration plots generated through 1000 bootstrap resamples, in addition to the Brier score and mean absolute error, to evaluate the consistency between the predicted and actual observed hearing loss outcomes. Risk stratification strategy was based on the optimal cutoff value (the hearing loss risk probability when the Youden index is maximum, Youden=sensitivity+specificity−1) [30]. Finally, the clinical utility of the final model was evaluated using decision curve analysis.

Ethical Considerations

The study was approved by the ethics committee of Capital Medical University, Beijing, China (Z2024SY071). Informed consent was obtained from participants during routine health examinations, and no additional consent was required for this secondary analysis. All data were anonymized and securely stored with access restricted to the research team. No financial or other compensation was provided, as the data were derived from routine public health services.

Results

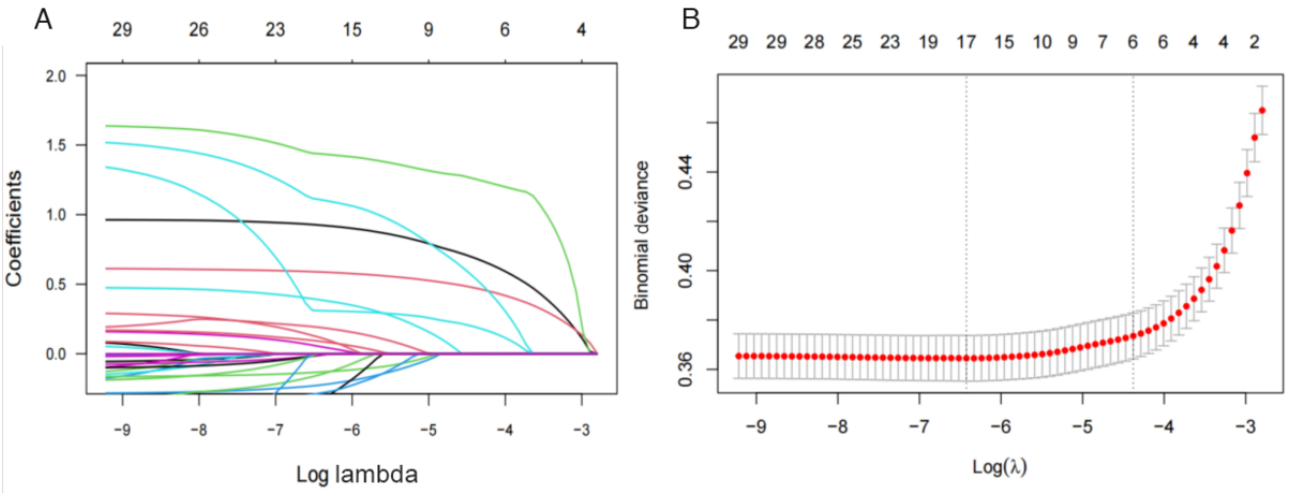
Population Characteristics

A total of 18,707 participants were included in the analysis in center A. Of these, 44.20% (8268/18,707) were male and 55.80% (10,439/18,707) were female, with a mean age of 72.89 (SD 7.15, range 65-99) years. A total of 6.29% (1177/18,707) participants were identified as having hearing loss, while 93.71% (17,530 /18,707) had normal hearing. The participants were then randomly divided into training (13,094/18,707, 70%) and internal validation sets (5613/18,707, 30%) in a 7:3 ratio. Additionally, 96.11% (11,008/11,454) and 95.40% (5268/5522) of participants were included from center B and center C, respectively, for external validation.

Model Development and Performance

The most significant features selected by LASSO from the training set were used to identify potential predictors for the model. The screening process is detailed in Figure 2. Initially, 33 variables were identified, guided by systematic reviews, meta-analyses, and expert clinical opinions. These factors encompass demographic characteristics, lifestyle factors, physiological measurements, and chronic diseases. The model's bias showed minimal variation within the interval (λ_{1se}, λ_{min}). When λ_{min} was set to 0.002 in the LASSO model, 17 variables exhibited nonzero coefficients, as shown in Table 1.

Figure 2. Feature selection using the least absolute shrinkage and selection operator (LASSO) regression. (A) The LASSO coefficient profile plot. Each colored line represents the coefficient path for 1 of the 33 candidate predictor variables as the regularization penalty (log lambda) increases. As the penalty grows, the coefficients of less important variables are shrunk toward 0. (B) Identification of the optimal penalization coefficient (λ) was achieved by 10-fold cross-validation and the minimum criterion. The left vertical line represents the mean square error, and the right vertical line represents the cross-validated mean square error within 1 SE of the minimum.



XSL•FO
RenderX

Table 1. Univariate analysis of potential predictors for hearing loss in the training set (n=13,094).

Variables	No hearing loss, n (%)	Hearing loss, n (%)	Chi-square (df)	*P* value
Sex			0.5 (1)	.49
Male	6832 (52.2)	453 (3.5)		
Female	5448 (41.6)	361 (2.8)		
Age (years)			677.3 (2)	<.001
65-70	6467 (49.4)	152 (1.2)		
71-80	3785 (28.9)	242 (1.8)		
≥81	2028 (15.5)	420 (3.2)		
Education level			33.3 (2)	<.001
Associate degree or above	2815 (21.5)	207 (1.6)		
Middle or high school	6395 (48.8)	343 (2.6)		
Primary school or below	3070 (23.4)	264 (2)		
Marital status			1.8 (1)	.19
Married	9470 (72.3)	611 (4.7)		
Other (Never married, divorced, or separated, widowed)	2810 (21.5)	203 (1.6)		
Waist circumference			0.3 (1)	.56
Normal	7737 (59.1)	504 (3.8)		
Abnormal	4543 (34.7)	310 (2.4)		
BMI (kg/m^2)			8.0 (2)	.02
18.5-29.9	11,708 (89.4)	759 (5.8)		
＜18.5	181 (1.4)	15 (0.1)		
＞30.0	391 (3)	40 (0.3)		
Physical function			483.1 (1)	<.001
Able to perform smoothly	12,203 (93.2)	739 (5.6)		
Unable to perform any task independently	75 (0.6)	77 (0.6)		
Exercise frequency			886.0 (3)	<.001
Everyday	7592 (58)	205 (1.6)		
More than once a week	1826 (13.9)	79 (0.6)		
Occasionally	970 (7.4)	75 (0.6)		
Not exercising	1892 (14.4)	455 (3.5)		
Dietary habits			57.2 (3)	<.001
Balanced diets	11,411 (87.1)	703 (5.4)		
Primarily vegetarian diet	527 (4)	59 (0.5)		
Primarily meat-based diet	182 (1.4)	21 (0.2)		
Preference for sugar, salt, and oil	160 (1.2)	31 (0.2)		
Smoking			11.9 (2)	.003
Never smoked	11,107 (84.8)	760 (5.8)		
Former smoker	360 (2.7)	25 (0.2)		
Current smoker	813 (6.2)	29 (0.2)		
Alcohol consumption frequency			13.5 (2)	.001
Never	11,138 (85.1)	768 (5.9)		
Occasionally	486 (3.7)	15 (0.1)		
Frequently or daily	656 (5)	31 (0.2)		

XSL•FO
RenderX

Variables	No hearing loss, n (%)	Hearing loss, n (%)	Chi-square (df)	*P* value
Hypertension			19.7 (1)	<.001
No	4785 (36.5)	253 (1.9)		
Yes	7495 (57.2)	561 (4.3)		
Chronic obstructive pulmonary disease			0.01 (1)	.90
No	12,115 (92.5)	804 (6.1)		
Yes	165 (1.3)	10 (0.1)		
Diabetes			0.8 (1)	.37
No	3730 (28.5)	260 (2)		
Yes	8550 (65.3)	554 (4.2)		
Stroke			2.1 (1)	.15
No	10,365 (79.2)	671 (5.1)		
Yes	1915 (14.6)	143 (1.1)		
Left systolic blood pressure (mm Hg)			5.4 (1)	.02
90-140	10,573 (80.7)	725 (5.5)		
＜90 or ＞140	1707 (13)	89 (0.7)		
Right systolic blood pressure (mm Hg)			5.7 (1)	.02
90-140	10,597 (80.9)	727 (5.6)		
＜90 or ＞140	1683 (12.9)	87 (0.7)		

First, univariate analysis was conducted on the 17 variables identified through LASSO regression to evaluate their independent associations with hearing loss. The detailed results of this analysis, including odds ratios and *P* values for all variables, are presented in Table 1. The results indicated that age, education level, marital status, BMI, physical function, exercise frequency, dietary habits, smoking, alcohol consumption frequency, hypertension, left systolic blood pressure, and right systolic blood pressure were significantly associated with hearing loss.

Subsequently, a forward stepwise regression approach selected 6 predictors—age, education, exercise frequency, physical function, dietary habits, and hypertension—which were then included in the multivariate logistic regression model (Table 2). A user-friendly nomogram was developed, as illustrated in Figure 3. The optimal cutoff value was 0.805 based on the maximum Youden index. Accordingly, the population was classified into low (predicted probability <80.5%) and high (predicted probability≥80.5%) hearing loss risk groups. A comparison of the model's predicted risk classifications against the observed outcomes is presented in Table 3. For example, in the training set, of the 814 individuals, the model correctly identified 639 (78.5%) with hearing loss (true positives) and of the 12,280 individuals, 10,099 (82.2%) with normal hearing (true negatives), corresponding to the sensitivity and specificity values reported.

In the training set, the model achieved an AUC of 0.828 (95% CI 0.813-0.842), sensitivity was 0.785 (639 out of 814 participants who actually developed hearing loss were correctly predicted as high-risk), specificity was 0.822 (10,099 of the 12,280 participants who did not develop hearing loss were correctly predicted as low-risk), and balanced accuracy was 0.804. In the internal validation set, AUC was 0.806 (95% CI 0.782-0.831), sensitivity was 0.774, specificity was 0.820, balanced accuracy was 0.797, and Matthews correlation coefficient was 0.350.

XSL•FO
RenderX

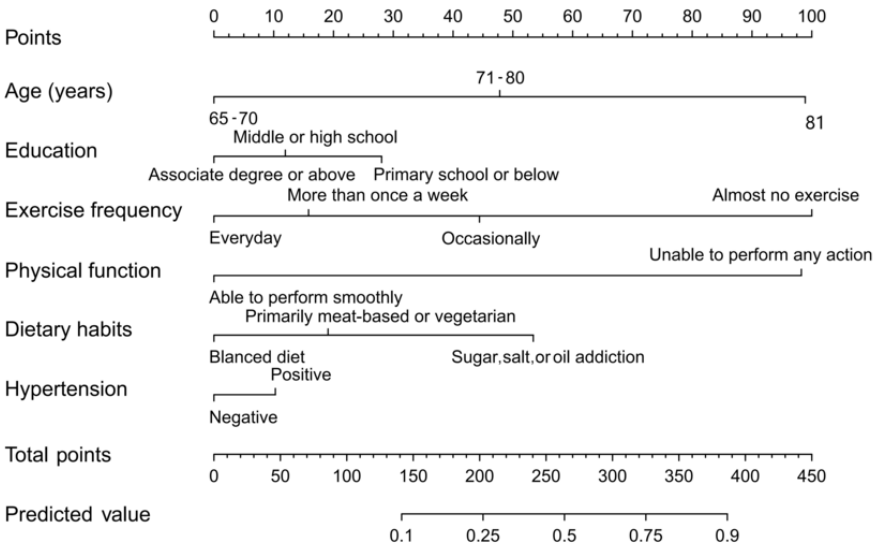
Table 2. Results of the multivariate logistic regression model for predicting hearing loss.

Predictor	β (SE)	OR[a] (95% CI)	*P* value
Intercept	−4.46 (0.131)	—[b]	<.001
Age (years)			
65-70 (reference)	—	—	—
71-80	1.15(0.108)	3.16 (2.56-3.91)	<.001
≥81	1.97 (0.108)	7.19 (5.81-8.89)	<.001
Education level			
Primary school or below (reference)	—	—	—
Middle or high school	−0.34 (0.095)	0.71 (0.59-0.86)	<.001
Associate degree or above	−0.49 (0.107)	0.61 (0.50-0.76)	<.001
Exercise frequency			
Every day (reference)	—	—	—
More than once a week	0.39 (0.133)	1.48 (1.14-1.91)	.003
Occasionally	0.75 (0.144)	2.13 (1.61-2.81)	<.001
Almost no exercise	1.84 (0.091)	6.3 (5.27-7.52)	<.001
Physical function			
Able to perform smoothly (reference)	—	—	—
Unable to perform any action	1.87 (0.176)	6.47 (4.58-9.14)	<.001
Dietary habits			
Balanced diets (reference)	—	—	—
Primarily meat-based or vegetarian	0.37 (0.135)	1.45 (1.11-1.89)	.006
Sugar, salt, and oil addiction	0.89 (0.23)	2.43 (1.55-3.82)	<.001
Hypertension			
Negative (reference)	—	—	—
Positive	0.26 (0.083)	1.29 (1.09-1.52)	.002

[a]OR: odds ratio.

[b]Not available.

Figure 3. Nomogram to predict the probability of hearing loss in community-dwelling older adults.



XSL•FO
RenderX

Table 3. Comparison of model-predicted risk stratification and observed hearing loss across training and validation cohorts.

Predicted risk group	Observed hearing loss, n	Observed no hearing loss, n	Row total, n
Training set			
Model-predicted high risk[a]	639	2181	2820
Model-predicted low risk[b]	175	10,099	10,274
Column total	814	12,280	13,094
Internal validation set			
Model-predicted high risk[a]	281	946	1227
Model-predicted low risk[b]	82	4304	4386
Column total	363	5350	5613
External validation set A			
Model-predicted high risk[a]	87	1285	1372
Model-predicted low risk[b]	39	9597	9636
Column total	126	10,882	11,008
External validation set B			
Model-predicted high risk[a]	148	791	939
Model-predicted low risk[b]	89	4240	4329
Column total	237	5031	5268

[a]The model-predicted low-risk group was defined as predicted probability <80.5%.

[b]The model-predicted high-risk group was defined as predicted probability ≥80.5%. Percentages are omitted because row percentages and column percentages represent different concepts. Absolute counts and all marginal totals are provided to allow direct calculation of any desired proportion.

External Validation

In external validation set A, the AUC was 0.720 (95% CI 0.670-0.771), with a sensitivity of 0.690, a specificity of 0.882, a balanced accuracy of 0.786, and a Matthews correlation coefficient of 0.204. In external validation set B, the model achieved an AUC of 0.747 (95% CI 0.712-0.782), with a sensitivity of 0.624, a specificity of 0.843, a balanced accuracy of 0.734, and a Matthews correlation coefficient of 0.270. These results highlight the model's good predictive performance for identifying hearing loss in older adults. The receiver operating characteristic curves are shown in Figure 4.

The nomogram's calibration plots are shown in Figure 5. Quantitative metrics supported the visual assessment of calibration. In the external validation set A, the Brier score was 0.125 and the mean absolute error was 0.002. In the external validation set B, these values were 0.221 and 0.003, respectively. Lower Brier scores and mean absolute error values suggest that the model provides probability estimates close to the true risk, helping clinicians avoid unnecessary interventions for patients with low risk and missed diagnoses in patients participants with high risk.

Figure 6 presents the decision curves for the nomogram. The high-risk threshold probability represents the predicted probability of hearing loss, above which clinicians would consider routine audiometric screening and early interventions, such as lifestyle modifications. In the decision curve analysis plots for both cohorts, the model (red line) consistently demonstrated greater net benefit than the default strategies (treat all or none) across a conservative overlapping threshold range of approximately 5% to 75%. Using the model to guide ARHL prediction and implement routine screening and early interventions provided an additional net benefit of up to 20% compared with alternative strategies that assume all patients are either severely or nonseverely affected.

XSL•FO
RenderX

Figure 4. Receiver operating characteristic curves of the prediction model for hearing loss in community-dwelling older adults in the training and validation sets.

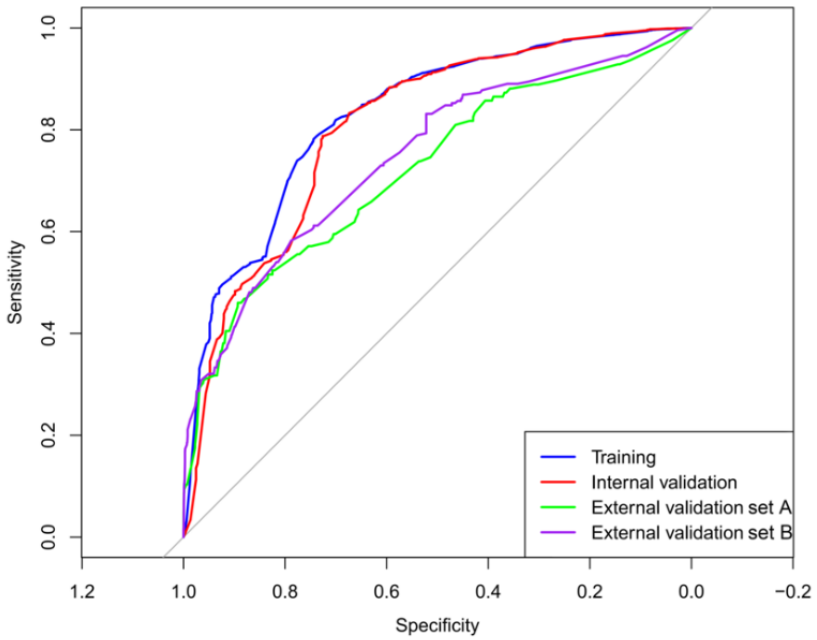
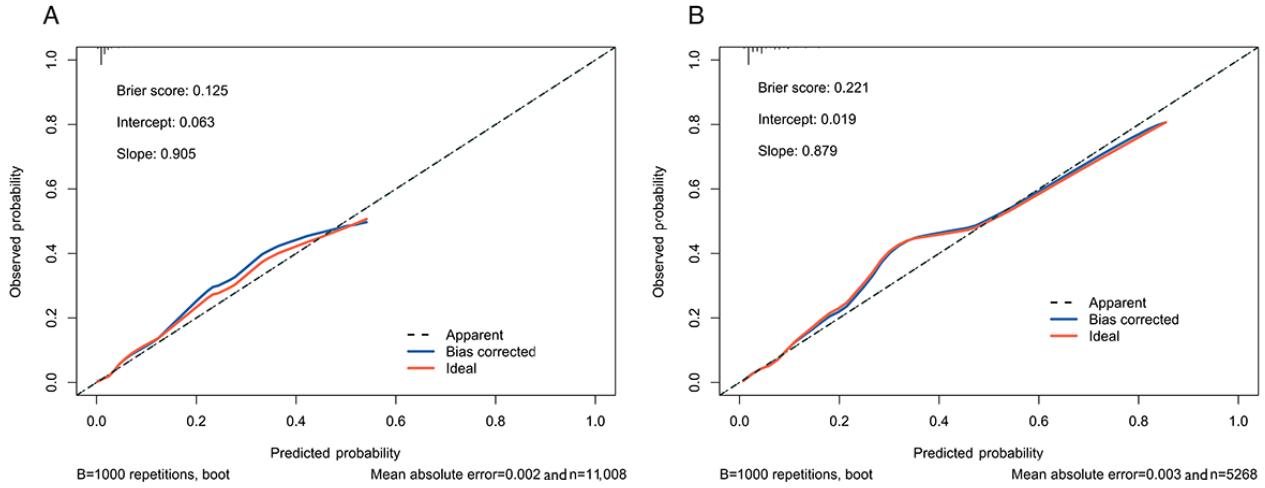
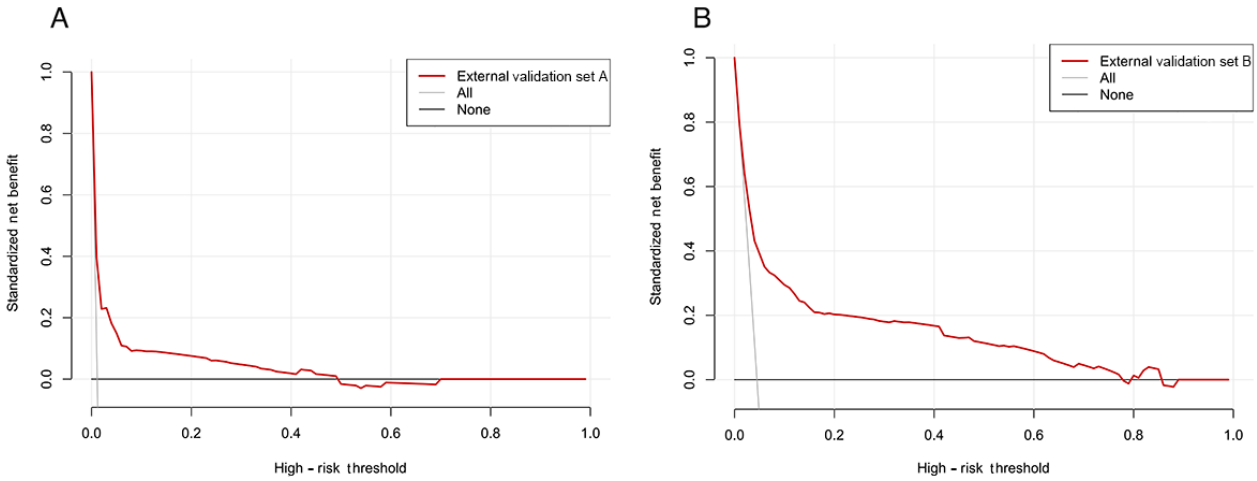


Figure 5. Calibration plots of the prediction model for hearing loss in community-dwelling older adults in external validations sets. (A) Calibration plot for external validation set A. (B) Calibration plot for external validation set B.



XSL•FO
RenderX

Figure 6. Decision curve analysis of the prediction model for hearing loss in community-dwelling older adults in external validation sets. (A) Decision curve analysis for external validation set A. (B) Decision curve analysis for external validation set B.

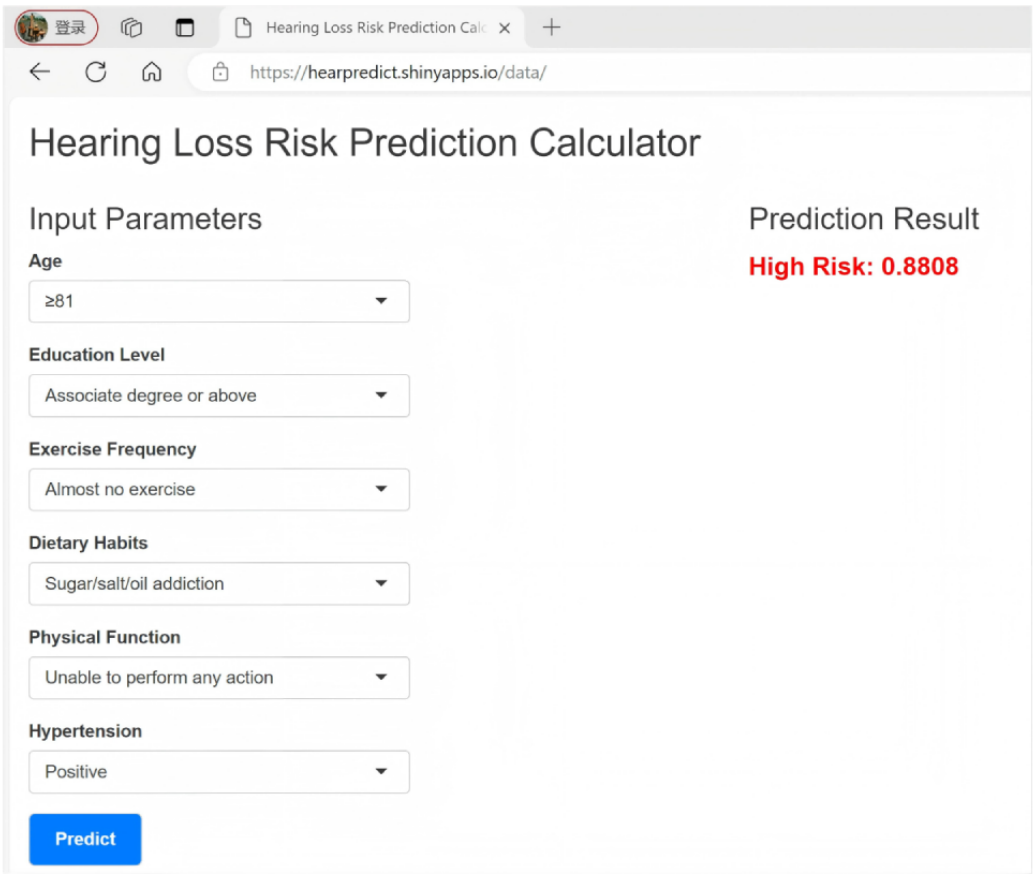


Visualization of Predictive Models: Development of a User-Friendly Web Page

To facilitate the clinical application of our nomogram, we developed an interactive and user-friendly web application using the R Shiny framework. The application features a simple, single-panel user interface where clinicians can input a patient's 6 predictor variables—age, education, exercise frequency, physical function, dietary habits, and hypertension—using a combination of numeric input fields and dropdown menus (Figure 7).

Upon clicking a *Predict* button, the tool instantly computes and displays the patient's predicted probability of hearing loss. The output is presented as a clear percentage, along with a corresponding risk stratification (*High risk* or *Low risk* based on the 80.5% cutoff) to aid in immediate clinical decision-making. The web application is publicly accessible and requires no software installation or programming knowledge, ensuring its broad utility for primary care providers [31].

Figure 7. Web page–based estimation of hearing loss likelihood in community-dwelling older adults.



XSL•FO
RenderX

Discussion

Principal Findings

The principal finding of this large, multicenter study is the development and external validation of a practical, 6-predictor nomogram for predicting hearing loss in community-dwelling older adults. Our model, which uses readily available data from routine health checkups (age, education, exercise frequency, physical function, dietary habits, and hypertension), demonstrated good discriminative ability (AUC>0.720 in all cohorts) and calibration.

A key clinical implication of this finding is the model's potential to shift primary care from reactive secondary screening to proactive primary prevention. The online tool, developed as a proof of concept, illustrates how this model could be integrated into clinical workflows. For example, if embedded within electronic medical record systems, it could provide real-time risk scores, creating a risk-stratified screening pathway. This would allow general practitioners to prioritize formal audiometry and targeted interventions for individuals at the highest risk, potentially improving the efficiency of health care resource allocation.

Comparison With Prior Work

Our risk prediction model incorporates demographic, lifestyle, and chronic disease factors, aligning broadly with previous studies. It is well established that age is strongly associated with hearing loss. Hearing decline along with aging is primarily attributed to progressive degeneration of cochlear sensory hair cells, stria vascularis, and spiral ganglion neurons, leading to impaired sound transmission [18,32]. Consistent with previous studies, we found a higher risk of hearing loss in individuals with lower education levels [33,34]. This may be attributed to its correlation with lower socioeconomic status [35], which can limit health care access and use, ultimately leading to poorer health status. Additionally, lower education levels are often associated with unhealthy lifestyle attributes (eg, smoking, alcohol consumption, and exercise habits) that were included in our model [36]. Interestingly, while recent studies have suggested a potential link between diabetes and hearing loss [37,38], our study did not identify a statistically significant association between the two. Instead, hypertension was included as a predictor in the model. It may contribute to hearing decline by reducing blood supply to the stria vascularis, a cochlear structure dependent on terminal arteries without collateral circulation [39,40].

Recent studies have focused on a prediction model for hearing loss. Ge et al [41] developed a logistic regression model for ARHL using 18 key factors based on 401 older adults. Their model achieved an AUC value of 0.777, and the sensitivity and specificity were 70.9% and 75.3%, respectively; however, due to limited sample size and absence of internal and external validation, further scrutiny is required regarding its predictive performance. Similarly, Wang et al [42] developed a predictive model comprising 22 features. While the model achieved an AUC of 0.864 and 0.844 during training and internal validation, the large number of variables and complex operation may hinder its clinical applicability and generalizability. Predictive models should strike a balance between performance and practicality, aiming to achieve the highest possible predictive accuracy with the fewest possible variables [43].

Strengths and Limitations

The primary strengths of this study include its large, multicenter sample, which enhances the robustness and generalizability of our findings, and a rigorous variable selection process. The resulting 6-predictor model is simple, and its variables are accessible from routine physical examinations without requiring specialized tests. Furthermore, the development of an online tool highlights its potential feasibility for real-world implementation. Finally, the selection of an interpretable *white-box* model (logistic regression) was a deliberate choice. While more complex *black-box* algorithms may offer marginal gains in accuracy, their lack of transparency is a significant barrier to clinical adoption. Our model achieves an optimal balance between its robust, externally validated performance and the essential need for interpretability, making it well suited for guiding real-world clinical decisions.

However, this study is not devoid of limitations. First, our study was conducted exclusively in primary health care centers within Beijing. While we attempted to mitigate this by validating across 3 distinct centers, the shared urban setting may influence the results by limiting the model's generalizability. Second, our model intentionally omits key risk factors such as genetic predisposition, long-term noise exposure, and ototoxic medication use. This was a necessary trade-off to create a practical tool based solely on variables routinely collected in health checkups. However, we acknowledge that the absence of these predictors may limit the model's accuracy in populations where such factors are highly prevalent. Third, hearing loss was assessed using the whispered voice test rather than pure-tone audiometry. To mitigate this, all professionals followed a standardized protocol. However, this method may influence the results by underestimating mild hearing loss and being subject to interexaminer variability. Fourth, several predictors (eg, physical function, exercise frequency, and dietary habits) were based on self-report, which may be prone to information bias and reduce predictive precision. Finally, as a retrospective cohort study, our analysis can suggest temporal relationships between exposures and outcomes but cannot establish definitive causal inferences due to potential confounding and selection biases.

Future Directions and Conclusions

Building on this work, several avenues for future research are recommended. The immediate priority is to validate the current nomogram in broader and more diverse geographic and demographic cohorts, using pure-tone audiometry as the gold standard outcome for a more precise assessment. Future prospective studies could also enhance the model's accuracy by incorporating key variables not available in our dataset, such as genetic predisposition, long-term noise exposure, and objective measures of physical activity. Finally, as the current online tool is a proof of concept, dedicated implementation research is required. This should address the practical challenges of integrating the tool into diverse electronic medical record

XSL•FO
RenderX

systems and evaluate its real-world impact on clinical decision-making, patient outcomes, and cost-effectiveness.

In conclusion, this study provides a validated, practical tool that shows promise for identifying older adults at high risk of hearing loss. If confirmed in future studies, this nomogram could become a valuable asset in the primary prevention of ARHL.

Acknowledgments

The authors would like to thank the study participants, as well as the nurses and general practitioners from the primary health care institutions, for their invaluable support. The authors gratefully acknowledge the financial support from the Beijing Municipal Social Science Foundation (23SRB007). The authors used the generative AI tool (Google Gemini), to assist with language editing and improve the text's readability. The authors take full responsibility for the final publication.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to patient privacy regulation but are available from the corresponding author on reasonable request.

Authors' Contributions

Conceptualization: JL, SJ, LS, and JEL
Data curation: QS, HW, and YZ
Formal analysis: JL and SJ
Funding acquisition: LS
Writing original draft: JL, SJ, LS, JEL, QS, MS, HW, and YZ
Writing–review & editing: SJ, LS, MS and JEL

Conflicts of Interest

None declared.

References

1. GBD 2015 Disease Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. Lancet 2016 Oct 08;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
2. World report on hearing. World Health Organization (WHO). URL: https://www.who.int/publications/i/item/9789240020481 [accessed 2025-05-29]
3. Long M, Zheng X, Bu X. China hearing health report. Social Science Academic Press. 2020. URL: https://www.pishu.cn/zxzx/xwdt/574203.shtml [accessed 2025-05-29]
4. Lin FR, Pike JR, Albert MS, Arnold M, Burgard S, Chisolm T, ACHIEVE Collaborative Research Group. Hearing intervention versus health education control to reduce cognitive decline in older adults with hearing loss in the USA (ACHIEVE): a multicentre, randomised controlled trial. Lancet 2023 Sep 02;402(10404):786-797. [doi: 10.1016/S0140-6736(23)01406-X] [Medline: 37478886]
5. Rutherford BR, Brewster K, Golub JS, Kim AH, Roose SP. Sensation and psychiatry: linking age-related hearing loss to late-life depression and cognitive decline. Am J Psychiatry 2018 Mar 01;175(3):215-224 [FREE Full text] [doi: 10.1176/appi.ajp.2017.17040423] [Medline: 29202654]
6. Loughrey DG, Kelly ME, Kelley GA, Brennan S, Lawlor BA. Association of age-related hearing loss with cognitive function, cognitive impairment, and dementia: a systematic review and meta-analysis. JAMA Otolaryngol Head Neck Surg 2018 Feb 01;144(2):115-126 [FREE Full text] [doi: 10.1001/jamaoto.2017.2513] [Medline: 29222544]
7. Jiam NT, Li C, Agrawal Y. Hearing loss and falls: a systematic review and meta-analysis. Laryngoscope 2016 Nov;126(11):2587-2596. [doi: 10.1002/lary.25927] [Medline: 27010669]
8. Yévenes-Briones H, Caballero FF, Struijk EA, Rey-Martinez J, Montes-Jovellar L, Graciani A, et al. Association between hearing loss and impaired physical function, frailty, and disability in older adults: a cross-sectional study. JAMA Otolaryngol Head Neck Surg 2021 Nov 01;147(11):951-958 [FREE Full text] [doi: 10.1001/jamaoto.2021.2399] [Medline: 34554203]
9. Lin FR, Niparko JK, Ferrucci L. Hearing loss prevalence in the United States. Arch Intern Med 2011 Nov 14;171(20):1851-1852 [FREE Full text] [doi: 10.1001/archinternmed.2011.506] [Medline: 22083573]
10. Agrawal Y, Platz EA, Niparko JK. Prevalence of hearing loss and differences by demographic characteristics among US adults: data from the National Health and Nutrition Examination Survey, 1999-2004. Arch Intern Med 2008 Jul 28;168(14):1522-1530. [doi: 10.1001/archinte.168.14.1522] [Medline: 18663164]

XSL•FO
RenderX

11. Contrera KJ, Wallhagen MI, Mamo SK, Oh ES, Lin FR. Hearing loss health care for older adults. J Am Board Fam Med 2016 May 11;29(3):394-403 [FREE Full text] [doi: 10.3122/jabfm.2016.03.150235] [Medline: 27170797]
12. Danhauer JL, Celani KE, Johnson CE. Use of a hearing and balance screening survey with local primary care physicians. Am J Audiol 2008 Jun;17(1):3-13. [doi: 10.1044/1059-0889(2008/002)] [Medline: 18519575]
13. Zuo JL, Lan YN, Sun W, Qiao YF, Li J, Shang YY. [Preliminary study on hearing screening for the elderly population undergoing physical examinations]. Zhonghua Er Bi Yan Hou Tou Jing Wai Ke Za Zhi 2025 Feb 07;60(2):166-171. [doi: 10.3760/cma.j.cn115330-20240620-00365] [Medline: 40010784]
14. US Preventive Services Task Force, Krist AH, Davidson KW, Mangione CM, Cabana M, Caughey AB, et al. Screening for hearing loss in older adults: US preventive services task force recommendation statement. JAMA 2021 Mar 23;325(12):1196-1201. [doi: 10.1001/jama.2021.2566] [Medline: 33755083]
15. GBD 2019 Hearing Loss Collaborators. Hearing loss prevalence and years lived with disability, 1990-2019: findings from the Global Burden of Disease Study 2019. Lancet 2021 Mar 13;397(10278):996-1009 [FREE Full text] [doi: 10.1016/S0140-6736(21)00516-X] [Medline: 33714390]
16. Bu XK. Primary prevention and treatment of otological diseases and hearing loss: WHO's latest program and China’s response and reflections. Chin Sci J Hear Speech Rehab 2014;2:40-41 [FREE Full text]
17. Sydlowski SA, Marinelli JP, Lohse CM, Carlson ML, Hearing Health Collaborative. Hearing health perceptions and literacy among primary healthcare providers in the united states: a national cross-sectional survey. Otol Neurotol 2022 Sep 01;43(8):894-899 [FREE Full text] [doi: 10.1097/MAO.0000000000003616] [Medline: 35900911]
18. Lin FR. Age-related hearing loss. N Engl J Med 2024 Apr 25;390(16):1505-1512. [doi: 10.1056/nejmcp2306778]
19. Lin FR, Thorpe R, Gordon-Salant S, Ferrucci L. Hearing loss prevalence and risk factors among older adults in the United States. J Gerontol A Biol Sci Med Sci 2011 May;66(5):582-590 [FREE Full text] [doi: 10.1093/gerona/glr002] [Medline: 21357188]
20. Tang D, Tran Y, Dawes P, Gopinath B. A narrative review of lifestyle risk factors and the role of oxidative stress in age-related hearing loss. Antioxidants (Basel) 2023 Apr 04;12(4):878 [FREE Full text] [doi: 10.3390/antiox12040878] [Medline: 37107253]
21. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015 Jan 07;350:g7594 [FREE Full text] [doi: 10.1136/bmj.g7594] [Medline: 25569120]
22. National essential public health service standards (3rd Edition). National Health Commission of the People's Republic of China. URL: https://www.nhc.gov.cn/ewebeditor/uploadfile/2017/04/20170417104506514.pdf#page=93.13 [accessed 2025-05-29]
23. Eekhof JA, de Bock GH, de Laat JA, Dap R, Schaapveld K, Springer M. The whispered voice: the best test for screening for hearing impairment in general practice? Br J Gen Pract 1996 Aug;46(409):473-474 [FREE Full text] [Medline: 8949327]
24. Pirozzo S, Papinczak T, Glasziou P. Whispered voice test for screening for hearing impairment in adults and children: systematic review. BMJ 2003 Oct 25;327(7421):967 [FREE Full text] [doi: 10.1136/bmj.327.7421.967] [Medline: 14576249]
25. Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019 Mar 30;38(7):1276-1296 [FREE Full text] [doi: 10.1002/sim.7992] [Medline: 30357870]
26. 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 [FREE Full text] [doi: 10.1016/s0895-4356(96)00236-3] [Medline: 8970487]
27. Wang J, Xu Y, Liu L, Wu W, Shen C, Huang H, et al. Comparison of LASSO and random forest models for predicting the risk of premature coronary artery disease. BMC Med Inform Decis Mak 2023 Dec 20;23(1):297 [FREE Full text] [doi: 10.1186/s12911-023-02407-w] [Medline: 38124036]
28. Kang J, Choi YJ, Kim I, Lee HS, Kim H, Baik SH, et al. LASSO-based machine learning algorithm for prediction of lymph node metastasis in T1 colorectal cancer. Cancer Res Treat 2021 Jul;53(3):773-783 [FREE Full text] [doi: 10.4143/crt.2020.974] [Medline: 33421980]
29. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, et al. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA 2017 Oct 10;318(14):1377-1384. [doi: 10.1001/jama.2017.12126] [Medline: 29049590]
30. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology 2005 Jan;16(1):73-81. [doi: 10.1097/01.ede.0000147512.81966.ba] [Medline: 15613948]
31. Li J. Nomogram prediction tool for hearing loss. shinyapps. URL: https://hearpredict.shinyapps.io/data/ [accessed 2025-05-29]
32. Ren HM, Ren J, Liu W. Recognition and control of the progression of age-related hearing loss. Rejuvenation Res 2013 Dec;16(6):475-486 [FREE Full text] [doi: 10.1089/rej.2013.1435] [Medline: 23915327]
33. Chou C, Beckles GL, Zhang X, Saaddine JB. Association of socioeconomic position with sensory impairment among US working-aged adults. Am J Public Health 2015 Jun;105(6):1262-1268. [doi: 10.2105/ajph.2014.302475]

XSL•FO
RenderX

34. Cruickshanks KJ, Dhar S, Dinces E, Fifer RC, Gonzalez F, Heiss G, et al. Hearing impairment prevalence and associated risk factors in the Hispanic community health study/study of Latinos. JAMA Otolaryngol Head Neck Surg 2015 Jul 01;141(7):641-648 [FREE Full text] [doi: 10.1001/jamaoto.2015.0889] [Medline: 26021283]
35. Emmett SD, Francis HW. The socioeconomic impact of hearing loss in U.S. adults. Otol Neurotol 2015 Mar;36(3):545-550 [FREE Full text] [doi: 10.1097/MAO.0000000000000562] [Medline: 25158616]
36. Cruickshanks KJ, Nondahl DM, Tweed TS, Wiley TL, Klein BE, Klein R, et al. Education, occupation, noise exposure history and the 10-yr cumulative incidence of hearing impairment in older adults. Hear Res 2010 Jun 01;264(1-2):3-9 [FREE Full text] [doi: 10.1016/j.heares.2009.10.008] [Medline: 19853647]
37. Gupta S, Eavey RD, Wang M, Curhan SG, Curhan GC. Type 2 diabetes and the risk of incident hearing loss. Diabetologia 2019 Feb 6;62(2):281-285 [FREE Full text] [doi: 10.1007/s00125-018-4766-0] [Medline: 30402776]
38. Akinpelu OV, Mujica-Mota M, Daniel SJ. Is type 2 diabetes mellitus associated with alterations in hearing? A systematic review and meta-analysis. Laryngoscope 2014 Mar 07;124(3):767-776. [doi: 10.1002/lary.24354] [Medline: 23945844]
39. Jin X, Xu X, Wang J, Liu X, Deng X, Xie H. Association between hypertension and hearing loss: a systemic review and meta-analysis. Front Neurol 2024 Jan 7;15:1470997 [FREE Full text] [doi: 10.3389/fneur.2024.1470997] [Medline: 39839866]
40. Li P, Pang K, Zhang R, Zhang L, Xie H. Prevalence and risk factors of hearing loss among the middle-aged and older population in China: a systematic review and meta-analysis. Eur Arch Otorhinolaryngol 2023 Nov 13;280(11):4723-4737. [doi: 10.1007/s00405-023-08109-3] [Medline: 37439927]
41. Ge J, Yan Y, Zhu Y, Cheng X, Li H, Sun X, et al. Development and validation of the screening tool for age-related hearing loss in the community based on the information platform. Eur Arch Otorhinolaryngol 2024 Jun 11;281(6):2893-2903 [FREE Full text] [doi: 10.1007/s00405-023-08389-9] [Medline: 38206390]
42. Wang Y, Ye C, Wang D, Li C, Wang S, Li J, et al. Construction and evaluation of a high-frequency hearing loss screening tool for community residents. Int J Environ Res Public Health 2021 Nov 23;18(23):12311 [FREE Full text] [doi: 10.3390/ijerph182312311] [Medline: 34886032]
43. Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. BMJ 2024 Sep 03;386:e078276 [FREE Full text] [doi: 10.1136/bmj-2023-078276] [Medline: 39227063]

Abbreviations

ARHL: age-related hearing loss
AUC: area under the receiver operating characteristic curve
LASSO: least absolute shrinkage and selection operator

Edited by N Cahill; submitted 23.07.25; peer-reviewed by P Sooful, P Usuemerai, O Mbanefo; comments to author 17.09.25; revised version received 15.11.25; accepted 16.11.25; published 10.12.25.

Please cite as:
Li J, Jin S, Sun L, Liu JE, Shen Q, Shang M, Wang H, Zhao Y
Predicting Age-Related Hearing Loss in Community-Dwelling Older Adults: Multicenter Retrospective Cohort Study
Interact J Med Res 2025;14:e81135
URL: https://www.i-jmr.org/2025/1/e81135
doi:10.2196/81135
PMID:

©Jing Li, Shuai Jin, Liu Sun, Jun-E Liu, Qiang Shen, Miao Shang, Hanting Wang, Yuanyuan Zhao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 10.12.2025. 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

Investigating the Association Between Mean Arterial Pressure on 28-Day Mortality Risk in Patients With Sepsis: Retrospective Cohort Study Based on the MIMIC-IV Database

Qimin Chen[1], BS; Wei Li[1], BS; Ying Wang[1], MD; Xianjun Chen[1], MD; Dehua He[1], BS; Ming Liu[1], BS; Jia Yuan[1], MD; Chuan Xiao[1], MD; Qing Li[1], BS; Lu Chen[2], MD; Feng Shen[1], PhD

[1]Department of Critical Care Medicine, The Affiliated Hospital of Guizhou Medical University, Beijing Road No. 8, Yunyan District, Guiyang, Guizhou Province, China

[2]Clinical Trials Centre, The Affiliated Hospital of Guizhou Medical University, Guiyang, China

Corresponding Author:
Feng Shen, PhD
Department of Critical Care Medicine, The Affiliated Hospital of Guizhou Medical University, Beijing Road No. 8, Yunyan District, Guiyang, Guizhou Province, China

Abstract

Background: Sepsis is a globally recognized health issue that continues to contribute significantly to mortality and morbidity in intensive care units (ICUs). The association between mean arterial pressure (MAP) and prognosis among patients with patients is yet to be demonstrated.

Objective: The aim of this study was to explore the association between MAP and 28-day mortality in ICU patients with sepsis using data from a large, multicenter database.

Methods: This is a retrospective cohort study. We extracted data of 35,010 patients with sepsis from the MIMIC-IV (Medical Information Mart for Intensive Care) database between 2008 and 2019, according to the Sepsis 3.0 diagnostic criteria. The MAP was calculated as the average of the highest and lowest readings within the first 24 hours of ICU admission, and patients were divided into 4 groups based on the mean MAP, using the quadruple classification approach. Other worst-case indications from the first 24 hours of ICU admission, such as vital signs, severity of illness scores, laboratory indicators, and therapies, were also gathered as baseline data. The independent effects of MAP on 28-day mortality were explored using binary logistic regression and a two-piecewise linear model, with MAP as the exposure and 28-day mortality as the outcome variables, respectively. To address the nonlinearity relationship, curve fitting and a threshold effect analysis were performed.

Results: A total of 34,981 patients with sepsis were included in the final analysis, the mean age was 66.67 (SD 16.01) years, and the 28-day mortality rate was 16.27% (5691/34,981). The generalized additive model and smoothed curve fitting found a U-shaped relationship between MAP and 28-day mortality in these patients. The recursive algorithm determined the low and high inflection points as 70 mm and 82 mm Hg, respectively. Our data demonstrated that MAP was negatively associated with 28-day mortality in the range of 34.05 mm Hg-69.34 mm Hg (odds ratio [OR] 0.93, 95% CI 0.92-0.94; P<.001); however, once the MAP exceeded 82 mm Hg, a positive association existed between MAP and 28-day mortality of patients with sepsis (OR 1.01; 95% CI 1.01-1.02, P=.002).

Conclusions: There is a U-shaped association between MAP and the probability of 28-day mortality in patients with sepsis. Both the lower and higher MAP were related with a higher risk of mortality in patients with sepsis. These patients have a decreased risk of mortality when their MAP remains between 70 and 82 mm Hg.

(Interact J Med Res 2025;14:e63291) doi:10.2196/63291

KEYWORDS

mean arterial pressure; 28-day mortality; sepsis; MIMIC-Ⅳ; retrospective study; Medical Information Mart for Intensive Care IV

Introduction

Sepsis is a life-threatening organ dysfunction caused by a dysregulated host response to infection [1-3]. While mortality from sepsis has decreased over time after age standardization, it remains high [2-6]. According to the Global Burden of Disease Study, 48.9 million cases of sepsis were reported worldwide in 2017, accounting for 11 million deaths [7]. Additionally, sepsis is the most expensive disease to treat in the United States, costing $23.7 billion annually and accounting for 6.2 percent of all hospital admissions [8]. In low-income nations with a

XSL•FO
RenderX

high sepsis burden, this cost could be significantly greater [9,10]. Sepsis has thus emerged as a serious public health concern with substantial global implications and economic cost.

Despite significant attempts to provide novel organ support strategies and identify the underlying etiology of sepsis, the mortality rate remains high [5,9]. Research indicates that early treatment of sepsis can improve prognosis [1,11]. Therefore, early detection of possible risk factors is crucial for patients with a poor prognosis [12]. Mean arterial pressure (MAP) is one of the most commonly used parameters for the evaluation of sepsis severity [13]. MAP is the driving pressure of tissue perfusion and plays a key role in maintaining the perfusion of tissue and organs. While autoregulation of regional perfusion may protect vital organs such as the brain or kidney from systemic hypotension, tissue perfusion becomes linearly dependent on arterial pressure below a certain MAP (approximately 60 mm Hg) [13]. However, the optimal range of MAP during sepsis resuscitation remains undetermined. The 2016 and 2021 SSC Surviving Sepsis Campaign guidelines recommend that MAP be maintained at ≥65 mm Hg in septic shock [14,15] for improving tissue perfusion and prognosis [14,15]. However, a study by Vincent et al [16] found that MAP did not have a linear relationship with ICU mortality, and several other studies have shown that increased MAP does not necessarily improve clinical outcomes in patients with sepsis or septic shock [13,17-22]. On the contrary, excessively high blood pressure may increase the incidence of adverse events [17].

Given that the appropriate blood pressure range is yet to be determined in patients with sepsis, and considering limitations in sample sizes, differences in study design, variability in covariate adjustments, and the heterogeneity of patient populations included in previous studies, it is necessary to study and identify the appropriate blood pressure range for patients with sepsis based on large databases. Therefore, our study aims to investigate the relationship between MAP and the 28-day risk of mortality in patients with sepsis using Medical Information Mart for Intensive Care (MIMIC-IV), a large sample of US sepsis databases, to identify the inflection point value associated with a lower 28-day risk of death in sepsis. This large sample size will provide more stable and reliable results, allowing us to gain a better understanding of the relationship between MAP and the 28-day risk of mortality in sepsis.

Methods

Study Design and Setting

This retrospective cohort study analyzed data from the MIMIC-IV database, including patients admitted to Beth Israel Deaconess Medical Center between 2008 and 2019.Our study complied with the RECORD (Research Reports Using Observational Routine Collection of Health Data) statement [23].

Participants

The study population comprised adult patients with sepsis identified through *ICD-9* (*99,591 - 99,592*) and *ICD-10* (*R652*, *R6520*, *R6521*) diagnostic codes. Inclusion criteria were patients aged ≥18 years and with a confirmed sepsis diagnosis. Patients lacking MAP data were excluded. Of the initial 377,207 records, 34,981 patients met the eligibility criteria and were included in the final analysis.

Data Collection and Variables

Data extraction was performed by certified researchers following standardized MIMIC-IV database procedures. The primary exposure variable, MAP, was calculated as the average of highest and lowest readings within 24 hours of ICU admission. While MAP values naturally fluctuate during the course of sepsis, we chose this approach to provide clinically actionable targets, acknowledging that a more complex time-varying analysis might capture additional nuances in the relationship between MAP and outcomes. MAP was analyzed both as a continuous variable and categorized into quartiles (Q1: 34.05 - 69.34 mm Hg; Q2: 69.34 - 74.94 mm Hg; Q3: 74.94 - 81.87 mm Hg; and Q4: 81.87 - 159.47 mm Hg). The primary outcome was 28-day all-cause mortality, recorded as a binary variable (1=death, 0=survival).

Covariates were selected based on clinical relevance and previous literature [3,24-26]: demographic characteristics (gender, age, and race), vital signs (heart rate, respiratory rate, and temperature), laboratory measurements (eg, lactate), disease severity indices such as Charlson Comorbidity Index, sequential organ failure assessment (SOFA) score, and acute physiology and chronic health evaluation (APACHE III score), therapeutic interventions (ie, mechanical ventilation, renal replacement therapy, corticosteroids, vasoactive drugs, immunoglobulin, and antibiotics).

Statistical Analysis

The analytical approach comprised three sequential steps. First, we constructed univariate and multivariate binary logistic regression models with progressive covariate adjustment: model 1 (unadjusted), model 2 (adjusted for demographics), and model 3 (fully adjusted for all covariates). Continuous variables were presented as mean (SD) or median (range), and categorical variables as frequencies (percentages). Second, we employed generalized additive models (GAM) with smooth curve fitting to evaluate potential nonlinear relationships between MAP and mortality. For identified nonlinear associations, we determined inflection points using recursive algorithms and constructed piecewise linear regression models. Third, we conducted sensitivity analyses using alternative outcome measures (30-day mortality), different MAP calculation methods (median values), and varying covariate selection approaches. All analyses were performed using R software (version 4.3.2; R Foundation for Statistical Computing) and Empower Stats software (X&Y Solutions, Inc. Boston, MA) [27], with statistical significance set at $P<.05$ (two-sided).

Missing Data Handling

Patients with missing values for any of the following key variables were excluded from the analysis: MAP measurements, age, SOFA score, lactate levels, and mortality outcomes. To assess potential selection bias, we compared the characteristics of included and excluded patients.

XSL•FO
RenderX

Ethical Considerations

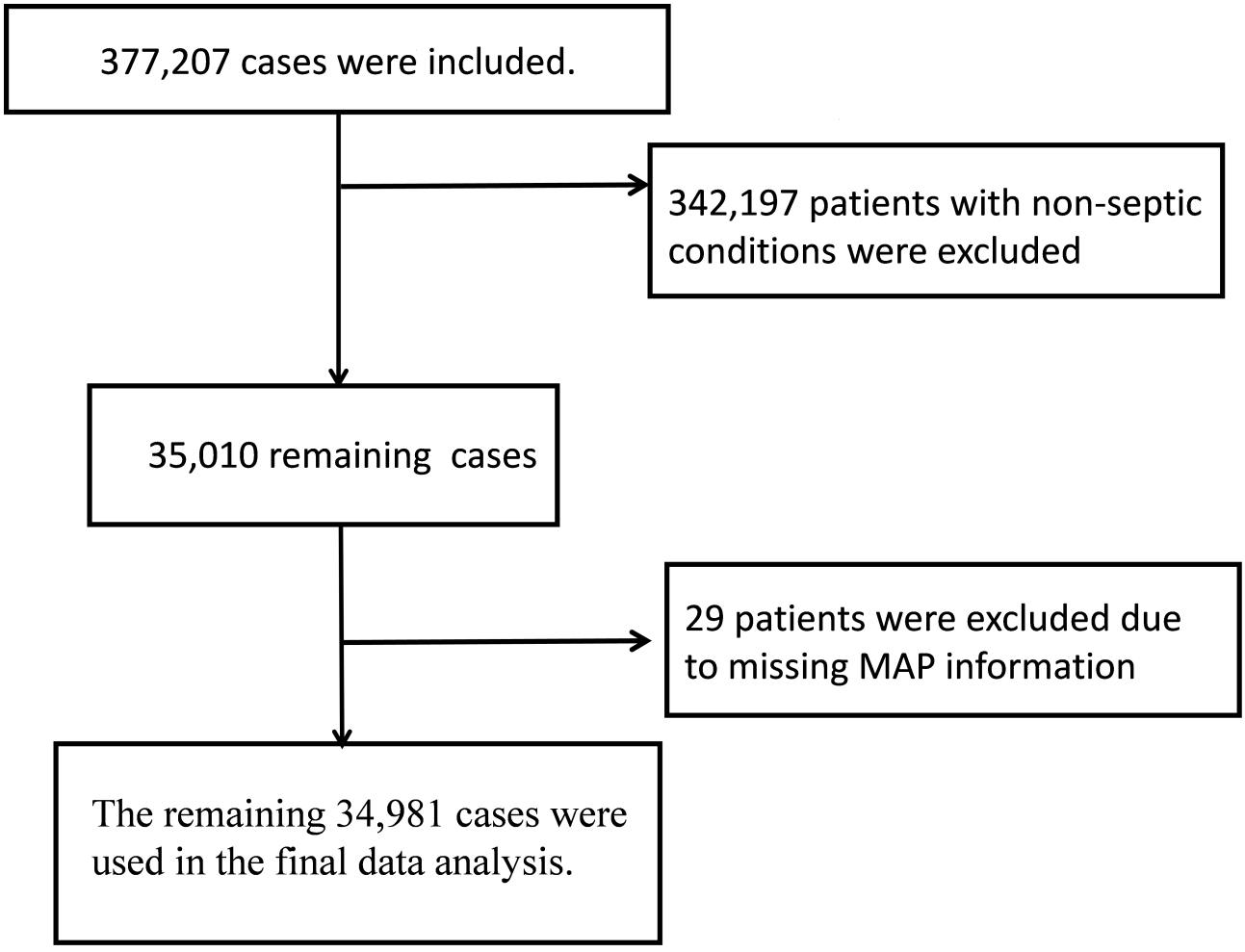
The study protocol received approval from the institutional review boards of Beth Israel Deaconess Medical Center (2001-P-001699/14) and Massachusetts Institute of Technology (0403000206). The data used in this study were obtained from the MIMIC-IV database, which is a freely accessible, deidentified public database [28]. Given the deidentified nature of the database, the requirement for informed consent was waived. Participants were not financially compensated, as the study used data from publicly accessible databases rather than data provided directly by participants.

Results

Study Population Screening

This study initially comprised 377,207 patients, of which 342,197 patients without sepsis and 29 patients lacking MAP data were excluded, with 34,981 cases remaining for final data analysis. The patient selection process is shown in the flowchart (Figure 1). Missing data were minimal (<5%, range 0 - 4.1%) across all variables. Given the low proportion of missing data, a complete case analysis was employed. The validity of this approach was confirmed by comparison of baseline characteristics between included and excluded patients, revealing no significant differences in key features.

Figure 1. The flowchart of study. MAP: mean arterial pressure.



Baseline Characteristics

Among the 34,981 included patients, 20,171 (57.66%) were men. The patients age was 66.67 (SD 16.01) years and the overall incidence of death within 28 days was 16.27% (5691/34,981). The patients' baseline characteristics are summarized in Table 1. Patients were divided into 4 groups based on MAP at admission. The distribution of intravenous immunoglobulin use did not differ significantly among the MAP subgroups (P=.09). Compared with the other three groups (Q2-Q4), patients in the Q1 group were older, had higher age, Charlson Comorbidity Index, SOFA scores, APACHE III scores, and lactate levels. They also had a higher frequency of renal replacement therapy, and use of norepinephrine, dopamine, dobutamine, cortisone, and antibiotics such as carbapenem, cephalosporin, and vancomycin. These characteristics, which are associated with poor prognosis, explain why patients in the Q1 group had the highest 28-day mortality rate. In contrast, as MAP increased from Q1 to Q4, age of the patient, severity of the disease, and the proportion of patients using relevant medications progressively decreased between groups. It may be a potential mechanism for the observed differences in 28-day mortality risk across MAP levels.

XSL•FO
RenderX

Table . Baseline characteristics of patients with sepsis in the MIMIC-IV database, 2008 - 2019.

Variables	Q1[a] (n=8745)	Q2[a] (n=8744)	Q3[a] (n=8745)	Q4[a] (n=8747)	*P* value
MAP[b] range (mm Hg)	(34.05 - 69.34)	(69.34 - 74.94)	(74.95 - 81.87)	(81.87 - 159.47)	_[c]
Age (years), mean (SD)	71.72 (15.36)	69.76 (15.09)	66.92 (15.33)	64.67 (16.39)	<.001
Males, n (%)	4687 (53.60)	5178 (59.22)	5180 (59.23)	5126 (58.60)	<.001
White population, n (%)	6350 (72.61)	6166 (70.52)	5860 (67.01)	5268 (60.23)	<.001
Charlson comorbidity index, mean (SD)	6.86 (2.83)	6.11 (2.87)	5.87 (2.92)	5.62 (3)	<.001
SOFA[e], mean (SD)	7.47 (4.02)	6.88 (3.79)	6.37 (3.56)	5.52 (3.22)	<.001
APACHE III[d], mean (SD)	81.43 (27.98)	72.11 (27.49)	67.21 (26.55)	64.24 (24.95)	<.001
Lactate (mmol/L), mean (SD)	3.55 (3.46)	3.14 (2.78)	3.00 (2.75)	2.84 (2.56)	<.001
Heart rate (bpm), mean (SD)	100.76 (25.11)	103.01 (24.08)	104.63 (24.15)	107.46 (24.23)	<.001
Respiratory rate (bpm), mean (SD)	26.67 (9.60)	26.57 (9.67)	26.49 (9.62)	27.08 (9.71)	<.001
Temperature (°C), mean (SD)	36.68 (1.28)	36.73 (1.30)	36.77 (1.25)	36.92 (1.25)	<.001
Dexamethasone, n (%)					<.001
No	7907 (90.42)	7917 (90.54)	7841 (89.66)	7497 (85.71)	
Yes	838 (9.58)	827 (9.46)	904 (10.34)	1250 (14.29)	
Methylprednisolone, n (%)					<.001
No	7297 (83.44)	7322 (83.74)	7315 (83.65)	6909 (78.99)	
Yes	1448 (16.56)	1422 (16.26)	1430 (16.35)	1838 (21.01)	
Cortisone, n (%)					.002
No	8526 (97.50)	8557 (97.86)	8577 (98.08)	8596 (98.27)	
Yes	219 (2.50)	187 (2.14)	168 (1.92)	151 (1.73)	
Norepinephrine, n (%)					<.001
No	4868 (55.67)	5520 (63.13)	6274 (71.74)	7240 (82.77)	
Yes	3877 (44.33)	3224 (36.87)	2471 (28.26)	1507 (17.23)	
Dopamine, n (%)					<.001
No	7783 (89)	8155 (93.26)	8223 (94.03)	8364 (95.62)	
Yes	962 (11)	589 (6.74)	522 (5.97)	383 (4.38)	
Dobutamine, n (%)					<.001
No	8241 (94.24)	8363 (95.64)	8377 (95.79)	8504 (97.22)	
Yes	504 (5.76)	381 (4.36)	368 (4.21)	243 (2.78)	
IVIG[f], n (%)					.09
No	8539 (97.64)	8538 (97.64)	8524 (97.47)	8495 (97.12)	
Yes	206 (2.36)	206 (2.36)	221 (2.53)	252 (2.88)	
MV[g], n (%)					<.001
No	5335 (61.01)	4517 (51.66)	4502 (51.48)	5222 (59.70)	

XSL•FO
RenderX

Variables	Q1[a] (n=8745)	Q2[a] (n=8744)	Q3[a] (n=8745)	Q4[a] (n=8747)	*P* value
Yes	3410 (38.99)	4227 (48.34)	4243 (48.52)	3525 (40.30)	
RRT[h], n (%)					<.001
No	7879 (90.10）	8232 (94.14）	8281 (94.69）	8225 (94.03）	
Yes	866 (9.90）	512 (5.86）	464 (5.31）	522 (5.97）	
Carbapenem, n (%)					<.001
No	6458 (73.85)	6883 (78.72)	7053 (80.65)	7138 (81.61)	
Yes	2287 (26.15)	1861 (21.28)	1692 (19.35)	1609 (18.39)	
Cephalosporin, n (%)					<.001
No	7892 (90.25)	8008 (91.58)	8007 (91.56)	8046 (91.99)	
Yes	853 (9.75)	736 (8.42)	738 (8.44)	701 (8.01)	
Penicillin, n (%)					<.001
No	3864 (44.19)	4381 (50.10)	4462 (51.02)	4200 (48.02)	
Yes	4881 (55.81)	4363 (49.90)	4283 (48.98)	4547 (51.98)	
Vancomycin, n (%)					<.001
No	1127 (12.89)	1662 (19.01)	1861 (21.28)	1909 (21.82)	
Yes	7618 (87.11)	7082 (80.99)	6884 (78.72)	6838 (78.18)	
28-day mortality, n (%)					<.001
No	6687 (76.47)	7452 (85.22)	7564 (86.50)	7587 (86.74)	
Yes	2058 (23.53)	1292 (14.78)	1181 (13.50)	1160 (13.26)	

[a]Patients were divided into 4 groups (Q1, Q2, Q3, and Q4) based on the mean MAP, using the quadruple classification approach.

[b]MAP: mean arterial pressure.

[c]Not applicable

[d]APACHE III: Acute Physiology and Chronic Health Evaluation III.

[e]SOFA: Sequential Organ Failure Assessment.

[f]IVIG: intravenous immunoglobulin.

[g]MV: mechanical ventilation.

[h]RRT: renal replacement therapy.

Relationship Between Mean Arterial Pressure and 28-Day Mortality

Univariate and Multivariate Analysis

We conducted univariate and multivariate analyses to investigate the relationship between MAP and 28-day mortality in patients with sepsis. Table 2 displays the results of the association between MAP and mortality in sepsis using various covariate adjustment strategies. In the nonadjusted model 1, each 1 mm Hg increase in MAP reduced the risk of 28-day mortality by 3% (odds ratio [OR] 0.97, 95% CI 0.971-0.974; *P*<.001). After adjusting for demographic factors such as gender, age, and White race, model 2 showed that the trend of OR was not altered (OR 0.97, 95% CI 0.97-0.98; *P*<.001). However, after adjusting for all covariables presented in Table 1, model 3 demonstrated that each 1 mm Hg increase in MAP resulted in a 1% decrease in the risk of 28 day-mortality (OR 0.99, 95% CI 0.99-1.00; *P*<.001). In model 3, the association between MAP and the probability of death in 28 days reduced to 1%, which was lower than the 3% decline seen in Models 1 and 2. This distinction was further explored and analyzed; we screened for covariates and discovered that SOFA score, age, and the Charlson comorbidity index had the greatest influence on the outcomes. The methodological rationale for this analysis was to include or exclude the effect of covariates on the regression coefficients of the main independent variables when constructing regression models. For example, in the unadjusted model, we included the main independent variable, MAP and found that the regression coefficient with the outcome was −0.0304. However, in the fully adjusted model, we added a number of covariates such as age, Charlson Comorbidity Index, SOFA score, and heart rate, the regression coefficient of MAP decreased to −0.0072 when these covariates were included (Multimedia Appendix 1). This implies that these factors “explain” or “mediate” the relationship between MAP and outcome to some extent. Specifically, age, Charlson index, and SOFA scores, which might reflect the patients’ baseline status and illness severity, contributed to the influence of MAP on outcome. This methodological aspect, namely the change in regression coefficients of the independent variables observed between the basic and full models, allows us to gain a deeper understanding of the complex link between independent factors and outcomes. It can help us uncover potential interactions or mediating effects that give a foundation

XSL•FO
RenderX

for understanding study findings. In addition, we performed a sensitivity analysis to verify the robustness of our findings by converting the MAP from a continuous variable into categorical variables (quartiles) and performing a trend test. The Q1 MAP was used as a reference; therefore, the same association was observed (P<.001) (Table 2).

Table . Relationship between mean arterial pressure and 28-day mortality.

Exposure	Model 1[a] (OR, 95% CI)	Model 2[b] (OR, 95% CI)	Model 3[c] (OR, 95% CI)	*P* value
MAP[d] (mm Hg)	0.97 (0.97-0.97)	0.97 (0.97-0.98)	0.99 (0.99-1.00)	<.001
MAP[e] quartiles (mm Hg)				
<65	1	1	1	
65 - 70	0.60 (0.55,-0.66)	0.61 (0.56 -0.67)	0.65 (0.59- 0.73)	<.001
70 - 80	0.41 (0.38-0.45)	0.42 (0.39- 0.46)	0.51 (0.46- 0.56)	<.001
80 - 85	0.38 (0.34-0.43)	0.40 (0.36- 0.45)	0.58 (0.51- 0.66)	<.001
≥85	0.38 (0.35- 0.42)	0.42 (0.38-0.47)	0.74 (0.65- 0.83)	<.001
P value for trend	<.001	<.001	<.001	_[f]

[a]Model 1 unadjusted.

[b]Model 2 adjusted for gender, age at admission, White race.

[c]Model 3 adjusted for gender, age at admission, White race.

[d]MAP as a continuous variable.

[e]MAP as a categorical variable.

[f]Not applicable.

Nonlinear Association Between MAP and 28-Day Mortality

We used smoothed curve fitting and GAM to evaluate the nonlinear relationship between MAP and 28-day mortality. After adjusting for all covariables listed in Table 1, our results demonstrated a U-shaped association between MAP and 28-day mortality with both low and high MAP associated with an increased risk of 28-day mortality (Figure 2). Furthermore, we employed a two-piecewise linear regression model and a recursive algorithm to determine the inflection points of MAP. The inflection points refer to points on the curve, where the curve transitions from falling to rising or rising to falling. Identifying these inflection points can help us better understand the complex relationship between MAP and 28-day mortality, as a single linear model may not fully capture this nonlinear relationship. We calculated the low and high inflection points at 70 and 82 mm Hg, respectively. When MAP was between 34.05 mm Hg and 70 mm Hg, the risk of 28-day mortality decreased by 7% for every 1 mm Hg increase (OR 0.93, 95% CI 0.92-0.94; P<.001). When the MAP was 70-82 mm Hg, there was no significant correlation (OR 1.01, 95% CI 0.99-1.02; P=.28). However, when MAP increased from 82 mm Hg to 159.47 mm Hg, the probability of 28-day mortality increased by 1% for every 1 mm Hg rise (OR 1.01, 95% CI 1.01-1.02, P=.002) (Table 3).

XSL•FO
RenderX

Figure 2. The nonlinear relationship between MAP and 28-day mortality risk. The solid line represents the smoothed curve fit while the dotted line represents 95% CI. MAP: mean arterial pressure.

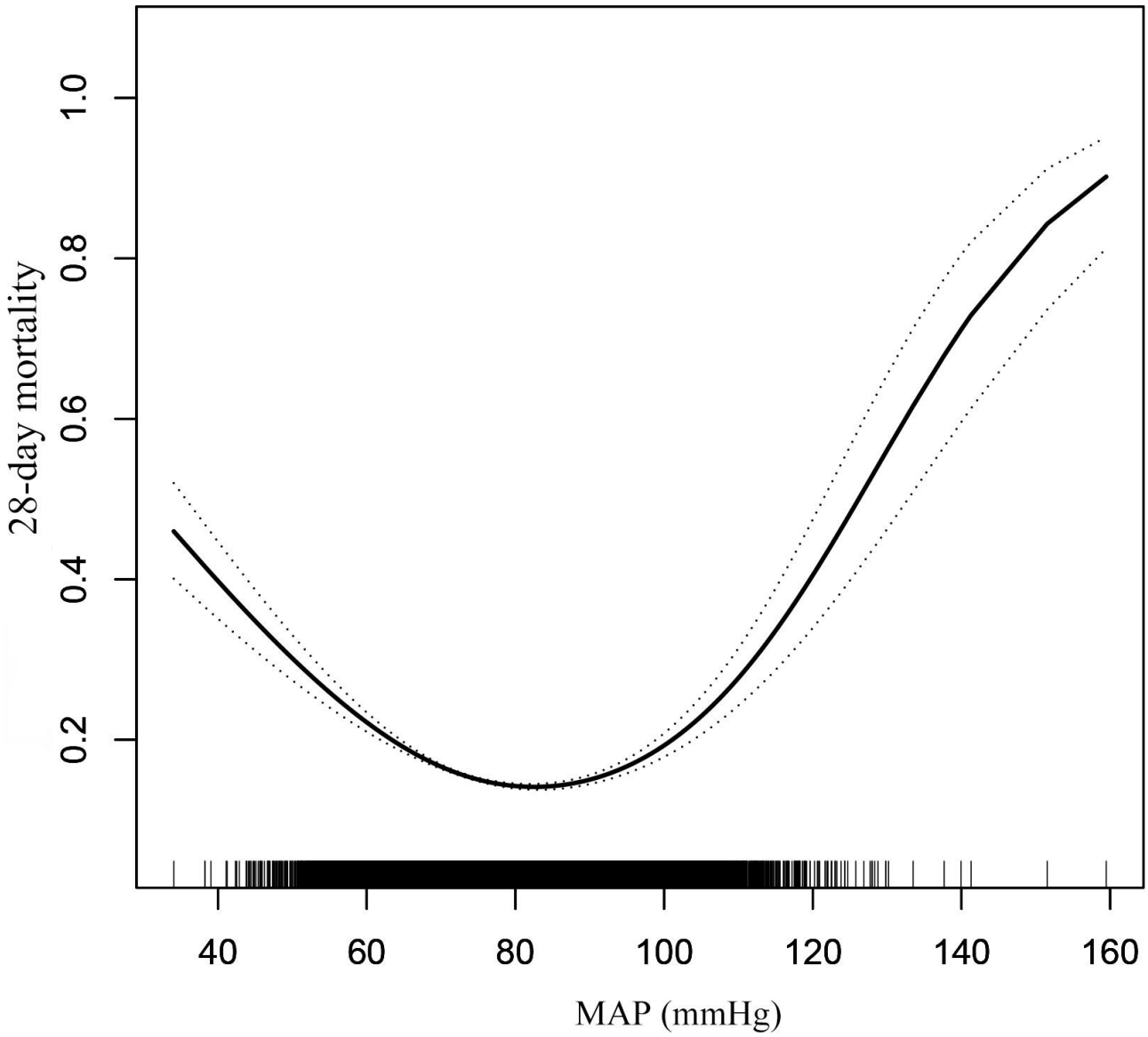


Table . Threshold effect analysis for the relationship between mean arterial pressure and 28-day mortality.

Outcomes	28-day mortality	
	OR (95%CI)	*P* value
Fitting model using logistic regression model	0.99 (0.99-1.00)	<.001
Fitting model using two-piecewise linear model		
Inflection point (mm Hg)		
<70	0.93 (0.92-0.94)	<.001
70 - 82	1.01 (0.99-1.02)	.28
>82	1.01 (1.01-1.02)	.002
Log-likelihood ratio test		<.001

Subgroup Analysis

The interaction analysis revealed significant effect modification by hypertension status (*P* for interaction=.03) and heart failure status (*P* for interaction=.03) on the association between MAP and 28-day mortality (Table 4). In hypertensive patients, each 1 mm Hg increase in MAP was associated with a slightly stronger protective effect against mortality (OR 0.993, 95% CI 0.992 - 0.994) compared to nonhypertensive patients (OR 0.984, 95% CI 0.982 - 0.986). Conversely, in patients with heart failure, the protective effect was marginally attenuated (OR 0.989, 95% CI 0.988 - 0.990) compared to those without heart

XSL•FO
RenderX

failure (OR 0.995, 95% CI 0.992 - 0.997). The use of vasopressors did not significantly modify the association between MAP and mortality (*P* for interaction=.14), with similar protective effects observed in both vasopressor users (OR 0.988, 95% CI 0.984 - 0.992) and nonusers (OR 0.991, 95% CI 0.990 - 0.992). These findings suggest that the optimal blood pressure management strategy may need to be tailored according to patients' comorbidity profile, particularly in those with hypertension or heart failure.

Table . Interaction analysis between mean blood pressure and treatment on 28-day mortality.

Model	28-day mortality, odds ratio (95% CI)	*P* value for interaction
Hypertension[a]		.03
No	0.984 (0.982-0.986)	
Yes	0.993 (0.992-0.994)	
Vasopressor[b]		.14
No	0.991 (0.990-0.992)	
Yes	0.988 (0.984-0.992)	
Heart failure[c]		.03
No	0.995 (0.992-0.997)	
Yes	0.989 (0.988-0.990)	

[a]Model Hypertension adjusted for: gender; age at admission; ethnicity; Charlson Comorbidity Index; dexamethasone; dopamine; dobutamine; IVIG, intravenous immunoglobulins; methylprednisolone; MV, mechanical ventilation; cortisone; SOFA, Sequential Organ Failure Assessment; glucocorticoids; cephalosporin; penicillin; heart rate; respiratory rate; vancomycin; body temperature and the interaction terms for following variables: dopamine, ethnicity, mechanical ventilation, SOFA.

[b]Model Vasopressor adjusted for: gender; age at admission; ethnicity; Charlson Comorbidity Index; dexamethasone; IVIG, intravenous immunoglobulins; methylprednisolone; MV, mechanical ventilation; cortisone; SOFA, Sequential Organ Failure Assessment; glucocorticoids; cephalosporin; penicillin; heart rate; respiratory rate; vancomycin; body temperature and the interaction terms for following variables: age at admission.

[c]Model Heart failure adjusted for: gender; age at admission; ethnicity; Charlson Comorbidity Index; dexamethasone; IVIG, intravenous immunoglobulins; methylprednisolone; MV, mechanical ventilation; cortisone; SOFA, Sequential Organ Failure Assessment; glucocorticoids; cephalosporin; penicillin; respiratory rate; vancomycin; body temperature and the interaction terms for following variables: age at admission, dexamethasone, IVIG, SOFA, glucocorticoids, cephalosporin, vancomycin, body temperature.

Sensitivity Analysis

Our sensitivity trials yielded results that confirmed the robustness of our findings (Multimedia Appendix 2). The regression coefficient for MAP was –0.0068 (95% CI –0.0123 to –0.0013) when the endpoint was 30-day mortality. When the median was used, the regression coefficient for MAP was –0.0065 (95% CI –0.0118 to –0.0012). Following the exclusion of individuals receiving medication for hypertension, the MAP regression coefficient was –0.0070 (95% CI –0.0128 to –0.0012). Other covariate selection methods produced results that were comparable. These results support our main finding, which confirms that the chance of patient mortality is inversely correlated with the MAP in all conditions.

Discussion

Main Findings

We investigated the relationship between MAP and the 28-day risk of mortality in patients with sepsis using data from a large, multicenter retrospective cohort study in the United States. We found a nonlinear association between MAP and 28-day mortality using a GAM and a two-piecewise linear regression. The findings revealed a U-shaped connection, indicating that both high and low MAP was associated with an increased risk of 28-day mortality in patients with sepsis. We discovered that each 1 mm Hg increase in MAP below 70 mm Hg resulted in a 7% decrease in the 28-day risk of mortality. When MAP exceeded 82 mm Hg, the 28-day mortality risk increased by 1% for each 1 mm Hg rise. This implies that maintaining MAP between 70-82 mm Hg may help improve the outcome in individuals with sepsis. While previous studies, including the work by Zhong et al [29] have explored the relationship between MAP and sepsis outcomes using MIMIC-III data, our study extends these findings in several important ways. First, we used the updated MIMIC-IV database with a larger sample size (34,981 vs. 14,031), providing more statistical power. Second, we employed more sophisticated statistical methods, including GAM models and two-piecewise linear regression, which allowed us to identify specific inflection points. Most importantly, our study identified a precise optimal MAP range (70-82 mm Hg) for clinical practice, whereas previous studies were unable to determine such specific thresholds. These differences explain why our findings provide more detailed and clinically applicable guidance for MAP management in patients with sepsis. There is no substantial relationship between MAP and 28-day mortality in this range, and organ damage from hypoperfusion or hyper-perfusion can be avoided. This finding provides clinicians with a reference range that can help guide blood pressure management in patients with sepsis. Further, more clinical studies are needed to validate these findings. However, this study provides a valuable reference for optimizing blood pressure management in patients with sepsis.

XSL•FO
RenderX

Most previous studies have explored the effect of low MAP on the outcome in patients with sepsis; our study further discovered a U-shaped relationship with specific inflection points. Our findings regarding the lower MAP threshold (70 mm Hg) are consistent with several previous studies. In a single-center retrospective cohort study of 1395 patients with severe sepsis or septic shock, []the 28-day mortality rate with an average MAP<65 mm Hg was 39.7%, which was significantly higher than other groups [30]. Similarly, in another retrospective cohort study based on the MIMIC III database by Cao et al. [31], which included 14,607 patients with sepsis, it was found that a lower MAP was significantly associated with higher 30-day mortality rates, with the lowest inflection point at 68.6 mm Hg.

However, our study differs from previous research in several important aspects. First, we used more sophisticated statistical methods including GAM models and two-piecewise linear regression, enabling us to identify both lower and upper thresholds. Although previous studies comparing the effects of higher and lower MAP on sepsis outcomes have yielded negative results, our findings uniquely identified 82 mm Hg as a critical upper threshold, above which there is an increased risk of death [17,18]. This increased risk could be due to secondary damage caused by excessive workload on organs and ischemia-reperfusion injury [31,32].

We speculate that the disparities between our findings and previous studies can be attributed to several factors: (1) the populations studied were not identical; (2) we provided more comprehensive covariate adjustment, whereas previous studies did not adjust for potential confounders; and (3) our analytical approach using the GAM model allowed us to detect nonlinear associations that might have been overlooked in previous studies. Further research is needed to determine whether these findings can be applied across populations.

Strengths and Limitations

This study has several key advantages. First, we used data from the updated MIMIC-IV cohort, which provides a larger sample size and, consequently greater statistical power. Second, we adjusted for a broader set of covariate indicators and focused on treatment strategies that were more closely related to patient outcomes. To better determine the true relationship between MAP and mortality, we used a GAM and two-piecewise linear models, which are advanced algorithms. Furthermore, we calculated the U-curve and identified two inflection points that provided a safe MAP range associated with lower mortality risk. Third, as this study used a large number of sensitivity analyses, the results are more robust. More informative, supporting evidence for MAP monitoring and clinical decision-making based on this metric is provided through this study.

Several limitations of this study should be noted. First, our findings are based solely on the MIMIC-IV database, which predominantly includes patients from a single geographic region and may not be representative of other populations. Given potential racial and ethnic differences in cardiovascular responses and outcomes, external validation of our findings in different 17 populations, particularly among Asian populations including Chinese patients, would be valuable. Such validation studies could help establish whether the optimal MAP range identified in our study is universally applicable or needs to be adjusted for different populations. Second, while we adjusted for measurable confounders, we could not account for unmeasurable confounding. Additionally, due to substantial missing data in lactate measurements, we were unable to conduct subgroup analyses based on lactate levels. This limitation prevents a deeper understanding of how the optimal MAP targets might differ among patients with varying degrees of tissue hypoperfusion, as indicated by lactate levels. Third, as this was an observational study, and therefore, inherently constrained, we could only observe relationships rather than evaluate causality.

Conclusion

There is a U-shaped association between MAP and the mortality risk in patients with sepsis. Both increases or decreases in MAP are linked to increased mortality risk. Our findings suggest that patients with sepsis have a lower risk of death when their MAP is maintained between 70-82 mm Hg.

Acknowledgments

The manuscript's English language quality was enhanced using AI-assisted proofreading tools, while maintaining the original content and meaning. This study was supported by the grants from National Natural Science Foundation of China [82160365]. This manuscript utilized AI tools exclusively for language enhancement and proofreading purposes. The intellectual content, research design, analyses, and conclusions were entirely conceived and executed by the authors.

Additionally, we confirm that one of our co-authors, LC, has completed the required human subjects training and obtained access to the MIMIC-IV database by signing the Data Use Agreement.

Data Availability

The data are available on the MIMIC-IV database website [33].

Authors' Contributions

Conceptualization: FS, QC
Data curation: YW
Formal analysis: LC, CX, DH, JY, ML, QC, QL, WL, XC

XSL•FO
RenderX

Methodology: QC
Supervision: FS
Writing – original draft: QC
Writing – review & editing: FS

Conflicts of Interest

None declared.

Multimedia Appendix 1
Baseline characteristics of patients with sepsis in the MIMIC-Ⅳ database, 2008-2019.
[DOCX File, 23 KB - i-jmr_v14i1e63291_app1.docx]

Multimedia Appendix 2
Relationship between mean arterial pressure and 28-day mortality.
[DOCX File, 17 KB - i-jmr_v14i1e63291_app2.docx]

References

1. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):801-810. [doi: 10.1001/jama.2016.0287] [Medline: 26903338]
2. Shankar-Hari M, Phillips GS, Levy ML, et al. Developing a new definition and assessing new clinical criteria for septic shock: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):775-787. [doi: 10.1001/jama.2016.0289] [Medline: 26903336]
3. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of clinical criteria for sepsis: for the third international consensus definitions for sepsis and septic shock (Sepsis-3). JAMA 2016 Feb 23;315(8):762-774. [doi: 10.1001/jama.2016.0288] [Medline: 26903335]
4. Vincent JL. Highlighting the huge global burden of sepsis. Anaesth Crit Care Pain Med 2020 Apr;39(2):171-172. [doi: 10.1016/j.accpm.2020.03.004] [Medline: 32145433]
5. Rhee C, Dantes R, Epstein L, et al. Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014. JAMA 2017 Oct 3;318(13):1241-1249. [doi: 10.1001/jama.2017.13836] [Medline: 28903154]
6. Prest J, Sathananthan M, Jeganathan N. Current trends in sepsis-related mortality in the United States. Crit Care Med 2021 Aug 1;49(8):1276-1284. [doi: 10.1097/CCM.0000000000005017] [Medline: 34261926]
7. 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. The Lancet 2020 Jan;395(10219):200-211. [doi: 10.1016/S0140-6736(19)32989-7] [Medline: 31911147]
8. Torio CM, Moore BJ. National inpatient hospital costs: the most expensive conditions by payer 2013. In: Healthcare Cost and Utilization Project (HCUP) Statistical Briefs: Agency for Healthcare Research and Quality; 2016. [Medline: 27359025]
9. Chiu C, Legrand M. Epidemiology of sepsis and septic shock. Curr Opin Anaesthesiol 2021 Apr 1;34(2):71-76. [doi: 10.1097/ACO.0000000000000958] [Medline: 33492864]
10. Vincent JL, Marshall JC, Namendys-Silva SA, et al. Assessment of the worldwide burden of critical illness: the intensive care over nations (ICON) audit. Lancet Respir Med 2014 May;2(5):380-386. [doi: 10.1016/S2213-2600(14)70061-X] [Medline: 24740011]
11. Wu M, Islam MM, Poly TN, Lin MC. Application of AI in sepsis: citation network analysis and evidence synthesis. Interact J Med Res 2024 Apr 15;13:e54490. [doi: 10.2196/54490] [Medline: 38621231]
12. Liang M, Xu Y, Ren X, Huang D, Jin M, Qiu Z. The U-shaped association between serum osmolality and 28-day mortality in patients with sepsis: a retrospective cohort study. Infection 2024 Oct;52(5):1931-1939. [doi: 10.1007/s15010-024-02256-3] [Medline: 38647828]
13. 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]
14. Rhodes A, Evans LE, Alhazzani W, et al. Surviving sepsis campaign: international guidelines for management of sepsis and septic shock: 2016. Intensive Care Med 2017 Mar;43(3):304-377. [doi: 10.1007/s00134-017-4683-6] [Medline: 28101605]
15. 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]
16. Vincent JL, Nielsen ND, Shapiro NI, et al. Mean arterial pressure and mortality in patients with distributive shock: a retrospective analysis of the MIMIC-III database. Ann Intensive Care 2018 Nov 8;8(1):107. [doi: 10.1186/s13613-018-0448-9] [Medline: 30411243]

XSL•FO
RenderX

17. 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]
18. Lamontagne F, Richards-Belle A, Thomas K, et al. Effect of reduced exposure to vasopressors on 90-day mortality in older critically ill patients with vasodilatory hypotension: a randomized clinical trial. JAMA 2020 Mar 10;323(10):938-949. [doi: 10.1001/jama.2020.0930] [Medline: 32049269]
19. Lamontagne F, Meade MO, Hébert PC, et al. Higher versus lower blood pressure targets for vasopressor therapy in shock: a multicentre pilot randomized controlled trial. Intensive Care Med 2016 Apr;42(4):542-550. [doi: 10.1007/s00134-016-4237-3] [Medline: 26891677]
20. Thooft A, Favory R, Salgado DR, et al. Effects of changes in arterial pressure on organ perfusion during septic shock. Crit Care Lond Engl 2011;15(5):R222. [doi: 10.1186/cc10462] [Medline: 21936903]
21. Hylands M, Moller MH, Asfar P, et al. A systematic review of vasopressor blood pressure targets in critically ill adults with hypotension. Can J Anesth/J Can Anesth 2017 Jul;64(7):703-715. [doi: 10.1007/s12630-017-0877-1]
22. 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]
23. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. PLoS Med 2015 Oct;12(10):e1001885. [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
24. Roggeveen LF, Guo T, Fleuren LM, et al. Right dose, right now: bedside, real-time, data-driven, and personalised antibiotic dosing in critically ill patients with sepsis or septic shock-a two-centre randomised clinical trial. Crit Care 2022 Sep 5;26(1):265. [doi: 10.1186/s13054-022-04098-7] [Medline: 36064438]
25. Maheshwari K, Nathanson BH, Munson SH, et al. The relationship between ICU hypotension and in-hospital mortality and morbidity in septic patients. Intensive Care Med 2018 Jun;44(6):857-867. [doi: 10.1007/s00134-018-5218-5] [Medline: 29872882]
26. Burstein B, Tabi M, Barsness GW, Bell MR, Kashani K, Jentzer JC. Association between mean arterial pressure during the first 24 hours and hospital mortality in patients with cardiogenic shock. Crit Care 2020 Aug 20;24(1):513. [doi: 10.1186/s13054-020-03217-6] [Medline: 32819421]
27. EmpowerStats: simple & powerful tool for epidemiology and bio-statistics that helps you analyze data and create publishable result tables, all without writing code. EmpowerStats. URL: http://www.empowerstats.com [accessed 2022-08-16]
28. Johnson AEW, Bulgarelli L, Shen L, et al. MIMIC-IV, a freely accessible electronic health record dataset. Sci Data 2023 Jan 3;10(1):1. [doi: 10.1038/s41597-022-01899-x] [Medline: 36596836]
29. Zhong X, Li H, Chen Q, et al. Association between different MAP levels and 30-day mortality in sepsis patients: a propensity-score-matched, retrospective cohort study. BMC Anesthesiol 2023 Apr 6;23(1):116. [doi: 10.1186/s12871-023-02047-7] [Medline: 37024806]
30. Lee GT, Hwang SY, Jo IJ, et al. Associations between mean arterial pressure and 28-day mortality according to the presence of hypertension or previous blood pressure level in critically ill sepsis patients. J Thorac Dis 2019 May;11(5):1980-1988. [doi: 10.21037/jtd.2019.04.108] [Medline: 31285891]
31. Cao B, Chen Q, Tang T, et al. Non-linear relationship between baseline mean arterial pressure and 30-day mortality in patients with sepsis: a retrospective cohort study based on the MIMIC-III database. Ann Transl Med 2022 Aug;10(16):872. [doi: 10.21037/atm-22-3457] [Medline: 36111019]
32. Cai J, Chen X, Liu X, et al. AMPK: The key to ischemia-reperfusion injury. J Cell Physiol 2022 Nov;237(11):4079-4096. [doi: 10.1002/jcp.30875] [Medline: 36134582]
33. Medical information mart for intensive care: freely available medical data for research. MIMIC. URL: https://physionet.org/content/mimiciv/2.1/ [accessed 2022-08-16]

Abbreviations

APACHE III: Acute Physiology and Chronic Health Evaluation III
ICU: intensive care units
MAP: mean arterial pressure
MIMIC-IV: Medical Information Mart for Intensive Care
SOFA: sequential Organ Failure Assessment

XSL•FO
RenderX

Edited by TDA Cardoso; submitted 15.06.24; peer-reviewed by Y Zhou, Z Zhang; revised version received 02.01.25; accepted 28.01.25; published 05.03.25.

Please cite as:
Chen Q, Li W, Wang Y, Chen X, He D, Liu M, Yuan J, Xiao C, Li Q, Chen L, Shen F
Investigating the Association Between Mean Arterial Pressure on 28-Day Mortality Risk in Patients With Sepsis: Retrospective Cohort Study Based on the MIMIC-IV Database
Interact J Med Res 2025;14:e63291
URL: https://www.i-jmr.org/2025/1/e63291
doi:10.2196/63291

© Qimin Chen, Wei Li, Ying Wang, Xianjun Chen, Dehua He, Ming Liu, Jia Yuan, Chuan Xiao, Qing Li, Lu Chen, Feng Shen. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 5.3.2025. 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

French Versions of 2 English Questionnaires on Problematic Digital Use Assessed by Adolescents and Their Parents: Cross-Cultural Linguistic Translation and Adaptation Study

Islam El Boudi[1], MEng; Mathilde Riant[1], PhD; Alexandre Bellier[1,2], MD, PhD; Nicolas Vuillerme[1,3], PhD

[1]Faculty of Medicine, AGEIS (Autonomie, Gérontologie, E-santé, Imagerie et Société), Grenoble Alpes University, Jean Roget Building, 3rd Floor, Grenoble, France

[2]Institut Universitaire de France, Paris, France

[3]Grenoble Clinical Investigation Center (Inserm CIC 1406), Grenoble Alpes University Hospital, Grenoble Alpes University, Grenoble, France

Corresponding Author:
Islam El Boudi, MEng
Faculty of Medicine, AGEIS (Autonomie, Gérontologie, E-santé, Imagerie et Société), Grenoble Alpes University, Jean Roget Building, 3rd Floor, Grenoble, France

Abstract

Background: The emergence of problematic digital use is increasingly alarming, affecting between 7% and 20% of the world's adolescent population. However, there is no validated questionnaire in French to measure this. Only a few questionnaires, either self-reported by adolescents or hetero-reported by parents, have been translated and validated in English.

Objective: This study aims to translate into French the Digital Addiction Scale for Children (DASC), which is self-reported by adolescents, and the Problematic Media Use Measure (PMUM), which is hetero-reported by parents of adolescents.

Methods: We used the "forward and backward" method to establish the translation and achieve cross-cultural adaptation with 8 parents and 8 adolescents aged between 12 and 17 years. There were three stages: (1) initial translation and synthesis or reconciliation of the translations phase; (2) back translation and expert committee phase; and (3) pretesting phase, during which 8 parents completed the PMUM questionnaire and 8 adolescents completed the DASC questionnaire.

Results: Despite slight variations in translation for both questionnaires, the translators quickly reached a consensus during the translation phase. The expert committee did not propose any other conceptual changes. In the final phase, the parents made no comments to improve the questions or the wording. Although some adolescents mentioned repetition between certain questions, they did not suggest any improvements to the DASC questionnaire in French.

Conclusions: Although the translated versions of the DASC and PMUM questionnaires provide a foundation for detecting problematic digital use, they require further validation studies to confirm their reliability and applicability in the French adolescent population.

(Interact J Med Res 2025;14:e55685) doi:10.2196/55685

KEYWORDS

internet addiction; screen time; questionnaire and surveys; translations; adolescents; mobile phone

Introduction

While smartphones make it easy to be entertained anywhere, computers, tablets, and video game consoles can also be a source of entertainment with negative impacts on health [1-3]. Meng et al [4] determined that 26% of the world's population have problematic use of smartphones and 16% have problematic use of social networks. Internet gaming disorder, recognized in the *DSM-5* (*Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*) [5], affects 6% of the population, rising to 22% of French adolescents [4].

Today, there is no consensus for defining digital addiction, but some researchers agree to say that, as a behavioral addiction, it is characterized by intensive use of all types of digital content with difficulty in self-regulation [4]. This implies difficulty in being aware, planning, controlling, and evaluating its limits regarding this use, both in time and in content [4,6]. We will refer to this as problematic digital use (PDU). Among the consequences of the PDU, we find that major exposure to television or video games has deleterious effects on students' attitudes toward school and is even linked to poorer academic performance [7,8]. In addition, this type of use is known to be linked to a lower sleep quantity and sleep quality [9,10], a more sedentary lifestyle [11], lower self-esteem, and more anxiety-depressive symptoms [10,12].

XSL•FO
RenderX

To estimate the prevalence of PDU, questionnaires are most often used in research, and some of them have already been validated in several countries and with several age profiles. To measure problematic smartphone use, we have identified 3 scales already validated in French. The first is the Internet Addiction Test (IAT)–smartphone [13] and is strongly inspired by the IAT developed by Young [14]. The second scale, Smartphone Addiction Scale short version [15], comprises 10 items and is based on the 6 symptoms present in the *DSM-5* to assess internet gaming disorder [5], such as loss of control or disruption of family. A final scale, Problematic Mobile Phone Use Questionnaire by Lopez-Fernandez et al [16], measures the same problem with 15 items divided into 3 dimensions: dangerous use, forbidden use, and dependent use of the smartphone. Three other questionnaires were translated and validated in French to assess problematic internet use. The IAT [17] has been validated on medical students, the Compulsive Internet Use Scale [18] is intended for adults, and the Generalized Problematic Internet Use Scale-2 [19] targets a general student population. Finally, the only questionnaire that measures problematic video game use in French is the Internet Gaming Disorder-20 [20], which has been validated with 20 items divided into 6 categories: salience, mood, tolerance, craving, conflict, and relapse.

To date, there exists no validated scale in French for quantifying digital addiction, and no measure has specifically addressed the adolescent population. However, adolescents are robust consumers of diverse digital content, perceiving it as a means to satisfy their autonomy needs and identity development, and to establish or maintain interpersonal connections [21,22]. While short-term benefits are highly valued by young people, the costs of intensive use are significant, leading to more risky behaviors, lower self-esteem, depressive symptoms, and suicidal ideation [23,24]. PDU can significantly alter the development of adolescents by causing neuronal dysfunctions in subcortical, frontal, and parietal areas involved in attention and behavioral control [21].

To measure PDU in adolescents, the English scale by Hawi et al [25], the Digital Addiction Scale for Children (DASC), is a promising tool. The scale is based on the 9 criteria recognized for diagnosing internet gaming disorder in the *DSM-5*. An existing alternative or complementary solution involves asking parents to assess their adolescents' PDU. Domoff et al [26] proposed a questionnaire in 2 versions: the final versions of the Problematic Media Use Measure (PMUM; 27 items) and the PMUM Short Form (9 items). The PMUM has been validated in English and considers the 9 criteria of the *DSM-5* for internet gaming disorder. Although translated into English and Arabic, it has never been translated into French.

Whereas most questionnaires measure specific device use (eg, smartphone) [13,15,16], this study aims to better measure PDU in the French adolescent population by linguistically translating 2 questionnaires: the DASC, administered to adolescents, and the PMUM, administered to parents of adolescents. The translation process was carried out using the "forward and backward" method.

Methods

Ethical Considerations

This study was approved by the Grenoble Alpes University Hospital Center and has received ethics approval from the South-East I Ethics Committee for the Protection of Individuals (approval 2022-A01943-40). The collected data were anonymized throughout the analysis process, in compliance with the MR003 methodology approved by the National Commission on Informatics and Liberty. For minor participants, consent to participate was obtained from the teenagers and their parents. Adult participants gave their own consent. Furthermore, no compensation was provided to the participants.

Cross-Cultural Adaptation Procedure With the "Forward and Backward" Method

Overview

We assessed the conceptual equivalence of the 2 questionnaires following the recommendations of Epstein et al [27] and used the most commonly adopted technique for cross-cultural research: the "forward and backward" method. Conceptual equivalence—evaluating whether items in the original language have a similar meaning in French—was emphasized by systematically analyzing item, semantic, idiomatic, operational, and experiential equivalence [28].

The forward and backward method is divided into 3 main phases of analysis, which we describe below.

Initial Translation and Synthesis or Reconciliation of the Translations Phase

We recruited 4 independent translators who are native French speakers and bilingual in English to translate the original French version. Two of the translators were familiar with the concept of PDU in their professional activities, while the other 2 did not work with this concept. From the 4 translations, the translators reconciled their differences to produce a single unified version.

Back Translation and Expert Committee Phase

The French version was back translated into English by a native English speaker to verify the accuracy of the translation and highlight any confusing wording. Then, 2 expert researchers familiar with the concept of PDU evaluated the equivalence of the questionnaire by comparing the back translated version with the original version. They proposed modifications to the items in French if necessary.

Pretesting Phase

We carried out a preliminary pilot study with adolescent participants and parents to assess the clarity, intelligibility, and acceptability of the French translations. The sampling was conducted using a convenience method through calls for participation in schools and libraries. A call for participation was shared during conference activities that we led about digital addiction. Participant criteria were specified to select specific settings such as age and region. Adolescents and parents responded to the call if they agreed to participate. Between December 2022 and June 2023, a total of 8 parents completed the PMUM questionnaire, and 8 adolescents aged between 12

XSL•FO
RenderX

and 17 years completed the DASC questionnaire. Accompanied by a student observer who cross-checked information to minimize interpretation biases, the principal researcher spoke with the participants, who gave their opinions on the questionnaires and assessed whether the questions were relevant to their age group. To categorize participants' feedback, we conducted a qualitative thematic analysis by manually coding their comments into four categories: (1) linguistic incomprehension, (2) linguistic ambiguity, (3) suggestions for modification, and (4) redundancies.

The protocol consisted of an individual interview with each participant, who completed the questionnaires. Each interview lasted a maximum of 30 minutes.

Instruments

The DASC questionnaire designed by Hawi et al [25] is composed of 25 statements distributed across 9 dimensions taken from the *DSM-5* video game addictive disorder criteria [5]: preoccupation in daily life, tolerance, withdrawal symptoms, mood changes, social and psychological consequences, relapses, physiological problems, deception regarding its use, and family disconnection. In the absence of the original online guideline that introduces the questionnaire to the participants, we took inspiration from the PMUM scale [26] by specifying that we want to capture adolescents' experiences with their screens (ie, television, video games, tablets, smartphones, portable game consoles, laptops, or computers). The adolescent must reflect on his or her experiences over the past year. The response scale ranges from 1=never to 5=always, and the total score can vary from 25 to 125.

The PMUM questionnaire is composed of 27 statements and was designed by Domoff et al [26], who based the instrument on the diagnostic criteria for online gaming addiction disorder in the *DSM-5* [5]. The guideline consists of asking parents how their child uses screens (ie, television, video games, tablets, smartphones, portable game consoles, laptops, or computers). For each statement, they select the option that is true for their child during the last year. For each question, the response scale is from 1=never to 5=always; the total average score on this scale can vary from 1 to 5.

Data Analysis

Data were analyzed using a qualitative approach with the forward and backward method, involving iterative interpretation of the interview responses. Descriptive statistics were conducted on the sample characteristics, such as age and gender. The qualitative analysis focused on identifying 4 key themes from the interviews.

Results

Characteristics of the Translators and the Participants

The age, gender, and specific cultural characteristics of the translators, the parent and adolescent participants are detailed in Table 1

XSL•FO
RenderX

Table . Age, gender, and cultural characteristics of the translators and the participants.

Step and participants	Age (years), mean (SD; range)	Gender	Cultural characteristics
Forward translation			
4 translators	25.5 (4.51; 20-31)	2 women and 2 men	• 2 native English speakers and 2 native French speakers • 2 were familiar with the concept of problematic smartphone use (eHealth and psychology) • 2 were not familiar with the concept (economics and education) • 2 were from the south and 2 from the north of France
Backward translation			
1 translator	28	1 woman	• Native English speaker • No clinical or medical expertise
Linguistic verification			
2 experts	27.5 (4.95; 24-31)	1 woman and 1 man	• 2 native French speakers
Final verification			
8 parents	43.1 (2.59; 40-48)	4 women and 4 men	• Good health and not experiencing any severe medical conditions • 2 from rural areas and 6 from urban areas in France
8 adolescents	13.8 (1.83; 12-17)	6 girls and 2 boys	• Good health or enduring any medical condition • 3 from rural areas and 5 from urban areas in France • 6 from middle school and 2 from high school

Forward Translation

Four translators translated the titles, dimensions, guidelines, response categories, and statements of the 2 questionnaires. The 4 proposals were grouped to arrive at a consensual version. Although the titles, response categories, and guidelines did not pose a translation problem, there was less consensus on some dimension names. This mainly concerns the "withdrawal" dimension present in the 2 questionnaires, which has ultimately been translated as "sevrage." The second, more problematic dimension to translate was that of the DASC scale: "displacement." In the context of behavioral addiction, displacement corresponds to the implementation of intense or compulsive displacement behaviors when an individual is in a stressful situation [29]. We opted for the "déplacement" term that is adapted and used in French [30]. Similarly, for the DASC scale, although "device" is generally translated in French as "appareil," we opted for "écrans" to better suit the context. For item 15, "enjoyable" offered multiple translation options, we translated it as "amusant" to adapt it to our context. For item 20, "I check my device…" had 4 different translations among the 4 translators as "to check" has a wide range of relevant translations. Therefore, we decided to translate it as "vérifier" even though it is not the most common word among all translation options, but as it is the most close translation from the original sentence without impacting the overall item meaning and conveying an intentional and repeating action. For item 25, "grades" was translated by "notes scolaires" as the raw translation "notes" would add too much ambiguity on the overall item.

For the PMUM questionnaire items, 23 of them were very similar between the 4 translations. Four items varied slightly for 1 or 2 translators. For instance, for item 10, the term "enjoys" offered multiple translation options; to achieve both semantic and idiomatic equivalence, we chose "apprécier," resulting in: "Il n'y a rien que mon enfant n'apprécie autant que les écrans." In 2 items, the term "amount of time" lacks a direct French equivalent. To preserve the operational and semantic meaning, we chose "la quantité de temps" for item 13 and "des périodes de plus en plus longues" for item 14. In item 22, the ambiguous term "sneaks" was resolved as "en cachette."

As for the DASC questionnaire items, 22 reached consensus, and 3 items varied slightly for 1 or 2 translators. Although "device" is typically rendered as "appareil," we selected "écrans" for better contextual equivalence. For item 15, to

XSL•FO
RenderX

capture the intended semantic nuance of "enjoyable," we chose "amusant." In item 20, given the wide semantic range of "to check," we opted for "vérifier" to best preserve the original meaning and indicate a repeated action. Finally, for item 25, "grades" was translated as "notes scolaires" to avoid ambiguity and ensure semantic clarity.

Backward Translation and Linguistic Verification

After the questionnaires were translated from the consensus version into English by the native English translator, we observed that all 21 out of 27 items of the PMUM questionnaire and all 25 items of the DASC questionnaire were accurately retranslated into English. For the PMUM questionnaire, some formulations or terms varied but did not modify the meaning of the items. One item's semantic meaning was strongly altered during the back translation: "My child attempts to use screen media for increasing amounts of time." was retranslated as "My child keeps using screens for longer periods." At the semantic level, this back translation does not reflect the intention of using the screen more and more—as covered by the initial item. Consequently, we modified the French item translation to better capture the original meaning of the item.

After these modifications, 2 researchers reread the latest French translation during a final professional translation step. No suggestions or conceptual modifications were proposed for the 2 questionnaires.

Final Verification

The final versions of the PMUM and DASC questionnaires were delivered to adolescents and parents in the pilot study. The 8 parents had no comments for improvement for any of the items or the statements, including linguistic incomprehension, linguistic ambiguity, suggestions for modification, or redundancies.

Although some adolescents noted repetitive phrasing (redundancies) across items addressing withdrawal symptoms—for example, items 12 and 21 use "frustrated" while items 3 and 8 use "upset," reflecting potential linguistic ambiguity—the expert team determined that these repetitions were essential for capturing subtle distinctions in behavior and experience, ensuring a comprehensive assessment of the construct, even if the phrasing occasionally feels redundant to adolescents, because "frustrated" and "upset" have close semantic meaning in French. No linguistic incomprehension was reported at this stage, and this feedback did not lead to suggestions for modification of the 25 items of the DASC questionnaire. The final translated versions are listed in Multimedia Appendix 1.

Discussion

Principal Findings

We effectively conducted a cross-cultural adaptation of 2 original questionnaires into French to measure PDU among adolescents. These questionnaires are intended to be filled out by adolescents themselves (DASC) or by their parents (PMUM). Although the DASC and PMUM questionnaires were originally validated in children aged 9 - 12 and 4 - 11 years, we applied these instruments to adolescents aged 12 - 17 years. One of the advantages of considering these 2 questionnaires is that, given that PDU in adolescents may manifest as extreme resistance to limits imposed by parents, leading to significant intrafamilial conflicts, questioning parents in addition to adolescents provides valuable information to better understand the extent of PDU. To make these questionnaires accessible in French, we translated them from English using the forward and backward method by following the 3 stages of translation of the method recommended by Epstein et al [27] and that has already been used for prior related French translations [31]. The conceptual equivalence of the French questionnaires was close to that of the English questionnaires. Regarding our results, neither parents nor adolescents encountered major difficulties in understanding the items, and experts recognized that the final translations were equivalent to the original questionnaires. Limitations include the small sample size, although the minimum number of participants for linguistic validation is not clearly defined [28], and the lack of validation. Further studies are required to validate these translations in larger populations and confirm their psychometric properties, particularly since adolescents reported that some items of the DASC questionnaire were redundant. Although other cross-cultural adaptations studies incorporated additional steps, such as confirmatory and exploratory factor analyses, to examine the instrument's internal structure and reliability in a simulated clinical setting [32-34], our study did not include these analyses. Yet, as the DASC questionnaire aims to assess PDU—a construct that is not directly observable and is composed of multiple dimensions such as tolerance, withdrawal symptoms, and overuse or physical and psychological consequences—future research should further investigate its structure validity. Additionally, researchers will need to assess the convergent validity of the questionnaires by showing that the scores obtained from the 2 translated questionnaires are positively correlated with other similar or closely related measures. These questionnaires can also be supplemented by objective and passive measurements derived directly from digital tools [35]. The last step will consist of evaluating the reliability of the questionnaires about their consistency and stability with longitudinal test-retest analyses.

These questionnaires, validated in French, will be useful in research to make intercultural comparisons in PDU demographics. Additionally, they can be used for professional purposes as part of medical or psychological care to detect PDU among adolescents.

Conclusions

This study enabled the effective cross-cultural adaptation of 2 questionnaires on the PDU of adolescents into French.

This study revealed that parents and adolescents easily understood the items, and experts confirmed the final translations' equivalence with the originals. Once validation of the questionnaires in French is completed, these 2 questionnaires can be used in French-speaking regions to screen or measure PDU among adolescent populations. Limitations of this research include the small sample size and the lack of psychometric validation, which are necessary to establish the reliability and validity of the translated questionnaires.

XSL•FO
RenderX

This study provides a questionnaire enabling adolescents to assess their own digital addiction (DASC), and another on parents' perceptions of their children's use (PMUM), which will enable the research community to better detect or assess the prevalence of the PDU in the French adolescent population.

Acknowledgments

This study would not have been possible without the support of Université Grenoble Alpes, Grenoble Alpes University Hospital Center, and Linksium (Society of Technology Transfer and Startup, Grenoble Alpes). This work was partially funded by MIAI (Multidisciplinary Institute in Artificial intelligence) Cluster (ANR-23-IACL-0006) and the French National Research Agency (Agence Nationale de la Recherche) within the framework of the "Investissements d'avenir" program (ANR-10-AIRT-05, 2012; ANR-15-IDEX-02, 2016). The funders played no role in this study's design, data collection, analysis, and interpretation of data, or writing of this paper. We used OpenAI's generative artificial intelligence tool, ChatGPT [36], to perform semantic reformulations to improve the overall quality of the English and to better understand the reviewers' expectations during the peer review process. The original ChatGPT transcripts are provided as Multimedia Appendix 2.

Data Availability

The datasets generated or analyzed during this study are not publicly available due to confidentiality considerations but are available from the corresponding author on reasonable request.

Authors' Contributions

IEB collected data, reviewed and assessed the results, wrote the original draft, wrote and reviewed the revised version of this paper, and acquired funding. MR collected, analyzed, and interpreted data; wrote the original draft; wrote and reviewed the revised version of this paper; and acquired funding. AB conceptualized and designed this study, reviewed and assessed the results, and wrote and reviewed the revised version of this paper. NV conceptualized and designed this study, supervised the project, reviewed and assessed the results, wrote and reviewed the revised version of this paper, and acquired funding. All authors reviewed and agreed to the published version of this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
DASC and PMUM questionnaires. DASC: Digital Addiction Scale for Children; PMUM: Problematic Media Use Measure.
[DOCX File, 27 KB - i-jmr_v14i1e55685_app1.docx]

Multimedia Appendix 2
ChatGPT conversation history.
[XLSX File, 12 KB - i-jmr_v14i1e55685_app2.xlsx]

References

1. 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]
2. Wacks Y, Weinstein AM. Excessive smartphone use is associated with health problems in adolescents and young adults. Front Psychiatry 2021;12:669042. [doi: 10.3389/fpsyt.2021.669042] [Medline: 34140904]
3. Zhang Y, Tian S, Zou D, Zhang H, Pan CW. Screen time and health issues in Chinese school-aged children and adolescents: a systematic review and meta-analysis. BMC Public Health 2022 Apr 22;22(1):810. [doi: 10.1186/s12889-022-13155-3] [Medline: 35459156]
4. Meng SQ, Cheng JL, Li YY, et al. Global prevalence of digital addiction in general population: a systematic review and meta-analysis. Clin Psychol Rev 2022 Mar;92:102128. [doi: 10.1016/j.cpr.2022.102128] [Medline: 35150965]
5. Diagnostic and Statistical Manual of Mental Disorders (DSM-5-TR). American Psychiatric Association. 2013. URL: https://www.psychiatry.org/psychiatrists/practice/dsm [accessed 2025-05-31]
6. LaRose R, Lin CA, Eastin MS. Unregulated internet usage: addiction, habit, or deficient self-regulation? Media Psychol 2003 Aug;5(3):225-253. [doi: 10.1207/S1532785XMEP0503_01]
7. Adelantado-Renau M, Moliner-Urdiales D, Cavero-Redondo I, Beltran-Valls MR, Martínez-Vizcaíno V, Álvarez-Bueno C. Association between screen media use and academic performance among children and adolescents: a systematic review and meta-analysis. JAMA Pediatr 2019 Nov 1;173(11):1058-1067. [doi: 10.1001/jamapediatrics.2019.3176] [Medline: 31545344]

XSL•FO
RenderX

8. Johnson JG, Cohen P, Kasen S, Brook JS. Extensive television viewing and the development of attention and learning difficulties during adolescence. Arch Pediatr Adolesc Med 2007 May;161(5):480-486. [doi: 10.1001/archpedi.161.5.480] [Medline: 17485625]
9. Chen YL, Gau SSF. Sleep problems and internet addiction among children and adolescents: a longitudinal study. J Sleep Res 2016 Aug;25(4):458-465. [doi: 10.1111/jsr.12388] [Medline: 26854132]
10. Ratan ZA, Parrish AM, Zaman SB, Alotaibi MS, Hosseinzadeh H. Smartphone addiction and associated health outcomes in adult populations: a systematic review. Int J Environ Res Public Health 2021 Nov 22;18(22):12257. [doi: 10.3390/ijerph182212257] [Medline: 34832011]
11. Zagalaz-Sánchez ML, Cachón-Zagalaz J, Sánchez-Zafra M, Lara-Sánchez A. Mini review of the use of the mobile phone and its repercussion in the deficit of physical activity. Front Psychol 2019;10:1307. [doi: 10.3389/fpsyg.2019.01307] [Medline: 31244720]
12. Prizant-Passal S, Shechner T, Aderka IM. Social anxiety and internet use – a meta-analysis: what do we know? What are we missing? Comput Human Behav 2016 Sep;62:221-229. [doi: 10.1016/j.chb.2016.04.003]
13. Barrault S, Durousseau F, Ballon N, Réveillère C, Brunault P. Smartphone addiction: French validation of the Internet Addiction Test-Smartphone Version (IAT-smartphone) and associated psychopathological features. Encephale 2019 Feb;45(1):53-59. [doi: 10.1016/j.encep.2017.12.002] [Medline: 29397925]
14. Young KS. Internet addiction: the emergence of a new clinical disorder. Cyberpsychol Behav 1998 Jan;1(3):237-244. [doi: 10.1089/cpb.1998.1.237]
15. Lopez-Fernandez O. Short version of the Smartphone Addiction Scale adapted to Spanish and French: towards a cross-cultural research in problematic mobile phone use. Addict Behav 2017 Jan;64:275-280. [doi: 10.1016/j.addbeh.2015.11.013] [Medline: 26685805]
16. Lopez-Fernandez O, Kuss DJ, Pontes HM, et al. Measurement invariance of the Short Version of the Problematic Mobile Phone Use Questionnaire (PMPUQ-SV) across eight languages. Int J Environ Res Public Health 2018 Jun 8;15(6):1213. [doi: 10.3390/ijerph15061213] [Medline: 29890709]
17. Khazaal Y, Billieux J, Thorens G, et al. French validation of the Internet Addiction Test. Cyberpsychol Behav 2008 Dec;11(6):703-706. [doi: 10.1089/cpb.2007.0249]
18. Khazaal Y, Chatton A, Horn A, et al. French validation of the Compulsive Internet Use Scale (CIUS). Psychiatr Q 2012 Dec;83(4):397-405. [doi: 10.1007/s11126-012-9210-x] [Medline: 22367623]
19. Laconi S, Kaliszewska-Czeremska K, Tricard N, Chabrol H, Kuss DJ. The Generalized Problematic Internet Use Scale-2 in a French sample: psychometric evaluation of the theoretical model. L'Encéphale 2018 Jun;44(3):192-199. [doi: 10.1016/j.encep.2017.09.001]
20. Plessis C, Altintas E, Romo L, Guerrien A. French validation of a scale evaluating internet gaming disorder: the Internet Gaming Disorder-20. Can J Psychiatry 2021 Dec;66(12):1051-1058. [doi: 10.1177/0706743721990796] [Medline: 33563027]
21. Marciano L, Ostroumova M, Schulz PJ, Camerini AL. Digital media use and adolescents' mental health during the COVID-19 pandemic: a systematic review and meta-analysis. Front Public Health 2021;9:793868. [doi: 10.3389/fpubh.2021.793868] [Medline: 35186872]
22. Uhls YT, Ellison NB, Subrahmanyam K. Benefits and costs of social media in adolescence. Pediatrics 2017 Nov;140(Suppl 2):S67-S70. [doi: 10.1542/peds.2016-1758E] [Medline: 29093035]
23. Vannucci A, Simpson EG, Gagnon S, Ohannessian CM. Social media use and risky behaviors in adolescents: a meta-analysis. J Adolesc 2020 Feb;79:258-274. [doi: 10.1016/j.adolescence.2020.01.014] [Medline: 32018149]
24. Vidal C, Lhaksampa T, Miller L, Platt R. Social media use and depression in adolescents: a scoping review. Int Rev Psychiatry 2020 May;32(3):235-253. [doi: 10.1080/09540261.2020.1720623] [Medline: 32065542]
25. Hawi NS, Samaha M, Griffiths MD. The Digital Addiction Scale for Children: development and validation. Cyberpsychol Behav Soc Netw 2019 Dec 1;22(12):771-778. [doi: 10.1089/cyber.2019.0132]
26. Domoff SE, Harrison K, Gearhardt AN, Gentile DA, Lumeng JC, Miller AL. Development and validation of the Problematic Media Use Measure: a parent report measure of screen media "addiction" in children. Psychol Pop Media Cult 2019 Jan;8(1):2-11. [doi: 10.1037/ppm0000163] [Medline: 30873299]
27. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol 2015 Apr;68(4):435-441. [doi: 10.1016/j.jclinepi.2014.11.021] [Medline: 25698408]
28. Harachi TW, Choi Y, Abbott RD, Catalano RF, Bliesner SL. Examining equivalence of concepts and measures in diverse samples. Prev Sci 2006 Dec;7(4):359-368. [doi: 10.1007/s11121-006-0039-0] [Medline: 16845592]
29. Anselme P. Abnormal patterns of displacement activities: a review and reinterpretation. Behav Processes 2008 Sep;79(1):48-58. [doi: 10.1016/j.beproc.2008.05.001]
30. Chabrol H, Callahan S. Mécanismes de Défense et Coping: Dunod; 2018. [doi: 10.3917/dunod.chabr.2018.01]
31. El Boudi I, Riant M, Bellier A, Vuillerme N. French versions of 4 English questionnaires on problematic smartphone use: cross-cultural linguistic translation and adaptation study. Interact J Med Res 2025 Feb 26;14:e53958. [doi: 10.2196/53958] [Medline: 40009839]
32. Bellier A, Chaffanjon P, Krupat E, Francois P, Labarère J. Cross-cultural adaptation of the 4-habits coding scheme into French to assess physician communication skills. PLoS ONE 2019;15(4):e0230672. [doi: 10.1371/journal.pone.0230672]

XSL•FO
RenderX

33. Zhao H, Rafik-Galea S, Fitriana M, Song TJ. Translation and psychometric evaluation of Smartphone Addiction Scale-Short Version (SAS-SV) among Chinese college students. PLoS ONE 2022;17(11):e0278092. [doi: 10.1371/journal.pone.0278092] [Medline: 36445890]
34. Adawi M, Bragazzi NL, Argumosa-Villar L, et al. Translation and validation of the nomophobia questionnaire in the Italian language: exploratory factor analysis. JMIR mHealth uHealth 2018 Jan 22;6(1):e24. [doi: 10.2196/mhealth.9186] [Medline: 29358161]
35. Ryding FC, Kuss DJ. Passive objective measures in the assessment of problematic smartphone use: a systematic review. Addict Behav Rep 2020 Jun;11(100257):100257. [doi: 10.1016/j.abrep.2020.100257] [Medline: 32467846]
36. ChatGPT (GPT-4). OpenAI. 2023. URL: https://openai.com/chatgpt [accessed 2025-03-27]

Abbreviations

DASC: Digital Addiction Scale for Children
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
IAT: Internet Addiction Test
PDU: problematic digital use
PMUM: Problematic Media Use Measure

Edited by TDA Cardoso; submitted 20.12.23; peer-reviewed by B Seebacher, G Padilla-Castillo; revised version received 10.04.25; accepted 22.04.25; published 09.07.25.

Please cite as:
El Boudi I, Riant M, Bellier A, Vuillerme N
French Versions of 2 English Questionnaires on Problematic Digital Use Assessed by Adolescents and Their Parents: Cross-Cultural Linguistic Translation and Adaptation Study
Interact J Med Res 2025;14:e55685
URL: https://www.i-jmr.org/2025/1/e55685
doi:10.2196/55685

© Islam El Boudi, Mathilde Riant, Alexandre Bellier, Nicolas Vuillerme. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 9.7.2025. 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 Correlates of Clinically Elevated Depressive Symptoms in a Nationwide Sample of Transgender, Nonbinary, and Gender Diverse Young Adults in the United States: Cross-Sectional Survey Study

Sari Reisner[1,2,3], ScD, MA; Yuxin Liu[4], MSc; Regina Tham[4], BS; Kaiden Kane[4], BS; S Wilson Cole[4,5], MA, MPH, PhD; Elizabeth R Boskey[6,7], LICSW, MSW, MPH, PhD; Sabra L Katz-Wise[8,9,10], PhD; Alex S Keuroghlian[3,11,12], MPH, MD; Rena Xu[4,7], MBA, MD

[1]Department of Epidemiology, University of Michigan School of Public Health, SPH-I 2649A, 1415 Washington Heights, Ann Arbor, MI, United States

[10]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[11]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[12]Psychiatry, Harvard Medical School, Boston, MA, United States

[2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, United States

[3]The Fenway Institute, Fenway Health, Boston, MA, United States

[4]Department of Urology, Boston Children's Hospital, Boston, MA, United States

[5]Query Research Consulting, Glen Burnie, MD, United States

[6]Division of Gynecology, Boston Children’s Hospital, Boston, MA, United States

[7]Department of Surgery, Harvard Medical School, Boston, MA, United States

[8]Division of Adolescent/Young Adult Medicine, Boston Children's Hospital, Boston, MA, United States

[9]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Sari Reisner, ScD, MA
Department of Epidemiology, University of Michigan School of Public Health, SPH-I 2649A, 1415 Washington Heights, Ann Arbor, MI, United States

Abstract

Background: In the United States, transgender, nonbinary, and gender diverse (TGD) young adults experience a higher risk of depression compared to their cisgender peers. Understanding factors associated with increased risk of depression within the TGD young adult population is important to guide clinical care as well as inform the development of interventions to reduce mental health disparities.

Objective: This exploratory study investigated the prevalence and correlates of positive screening for depressive symptoms among TGD young adults to inform the design, development, and implementation of national interventions aimed at improving mental health in this at-risk population.

Methods: In August 2022, a cross-sectional, nationwide online survey was conducted among TGD young adults aged 18 - 25 (N=104) in the United States. Measures included sociodemographic variables, family characteristics, mental health care utilization, and the two-item Patient Health Questionnaire-2 (PHQ-2) screener for depression. Poisson regression models with robust variance estimation were fitted to estimate adjusted prevalence ratios (aPR) and 95% CI for correlates of PHQ-2 depression (score ≥3).

Results: The study sample had a mean age of 22 (SD 2) years; 48/104 (46%) individuals identified as Black, Indigenous, or other People of Color, and 69/104 (66%) were nonbinary. Overall, 44 (42%) individuals screened positive for depression using PHQ-2. In a multivariable model adjusted for age, race and ethnicity, US census region, and health insurance status, factors associated with increased depression prevalence using PHQ-2 included low versus high family support (aPR 1.54, 95% CI 1.05 - 2.27) and identifying with a nonChristian religion versus being unaffiliated (aPR 1.66, 95% CI 1.04 - 2.63). Factors associated with reduced depression prevalence included living in a rural versus suburban area (aPR 0.48, 95% CI 0.26 - 0.92) and receiving mental health therapy versus not (aPR 0.71, 95% CI 0.53 - 0.97).

Conclusions: The high prevalence of depressive symptoms among TGD young adults in this study sample highlights the need for comprehensive mental health evaluation and support in this population. Depression risk is increased among certain subgroups, such as those with low family support. These findings are valuable in informing the development of interventions that aim to improve mental health outcomes among TGD young people.

XSL•FO
RenderX

(Interact J Med Res 2025;14:e66630) doi:10.2196/66630

KEYWORDS

transgender; depression; preventive screening; young adult; LGBTQ+; nonbinary; gender minority; gender diverse; mental health; prevalence; cross-sectional; survey; questionnaire; nationwide; USA; United States; North America

Introduction

In the United States, major depression is a leading cause of morbidity and mortality, affecting an estimated 21 million adults aged ≥18 years (8.3% of all US adults) [1,2]. It is most prevalent among young adults aged 18 - 25 (18.6%) [3]. The peak incidence of depression occurs between ages 12 - 25 years, and 1 in 5 adolescents (20.1%) aged 12 - 17 in the US has had at least one major depressive episode, highlighting the importance of early detection and intervention efforts [3,4]. The US Preventive Services Task Force recommends routine preventive screening for depression among the general adult population aged ≥18 years in primary care settings, using tools such as the Patient Health Questionnaire (PHQ) to identify individuals who should be further evaluated for depression and facilitate appropriate referrals to mental health services [5-7].

Transgender, nonbinary, and gender diverse (TGD) young adults in the United States have a two- to four-fold higher risk of major depression [8] and are more likely to screen positive for clinically elevated depressive symptoms relative to their cisgender counterparts [9]. For example, in a national random sample of 65,231 college students drawn from 71 US college campuses, the prevalence of past 2-week depression using PHQ-9 was 57.8% for TGD young adults compared to 28.4% for cisgender young adults [9]. Other studies report similarly high prevalence rates of past 2-week elevated depressive symptoms, which vary depending on the screening instrument used and specific age range sampled: 56.7% (PHQ-9 depression in a clinical TGD sample aged 13 - 20 y) [10], 57.9% (two-item Patient Health Questionnaire, PHQ-2 depression in a representative school-based sample of TGD 9th and 11th graders in Minnesota) [11], and 75% (PHQ-2 depression in nonprobability online sample of 5753 TGD young people aged 13 - 24 y) [12]. Additional research is needed to understand factors associated with screening positive for depression among TGD young people, particularly using brief instruments appropriate for universal screening in clinical settings. Identifying TGD young people at increased risk for depression may facilitate more effective clinical care and inform the development of interventions to address mental health disparities.

Sociodemographic differences in depression prevalence based on gender identity and race or ethnicity have been inconsistent across studies among TGD young people [13], although nonbinary individuals reported consistently high rates of depression [9]. Other sociodemographic factors, such as geographic region and urbanity, are also important to consider, particularly in the context of legislation restricting elements of care for TGD populations. Several studies document the adverse negative health impacts of antitransgender legislation, even among TGD people who do not live in states with active or pending legislation [14]. While family support is protective for TGD mental health [15,16], other family-related factors, such as religiosity, necessitate further exploration [11], particularly since TGD young people have reported mixed quantitative findings on religiosity and mental health [17], and attitudes toward TGD people may differ by religious affiliation [18,19]. Additionally, both current utilization of mental health services and a history of seeking resources or information about mental health services may be relevant for future intervention implementation for TGD young adults. Consistent with a transformative paradigm in mental health that emphasizes early prevention and intervention of emerging mental health disparities in individuals aged 12 - 25 years [20], patterns of adolescent help-seeking may persist into young adulthood. Some TGD young people may have sought mental health information, while others may not have, highlighting the need to better understand whether help-seeking behaviors differ between those with and without depression. This knowledge can inform targeted intervention strategies to reach those who may need help.

This exploratory study sought to investigate the prevalence and correlates of depression using PHQ-2 in a nationwide US sample of TGD young adults to understand factors associated with increased risk of depression among TGD young people and to inform the development of interventions aimed at reducing mental health disparities in this population.

Methods

Setting

In August 2022, a cross-sectional survey was conducted with a nationwide US sample of young adults recruited and enrolled through Prolific, a survey platform that hosts a diverse panel of adults aged ≥18 years. The survey was conducted as part of TransHealthGUIDE, a National Institutes of Health (NIH)-funded project (U01MH136558) aimed at designing and implementing strategies to address mental health disparities among TGD young adults. Early in-depth key informant interviews with TGD young adults, caregivers, and health care providers identified the need to better understand the context of depression among TGD young adults. Specifically, the interviews highlighted the need for further information on the characteristics of TGD young adults with and without depression, the role of family, mental health service utilization, peer support, and recalled help-seeking behaviors in adolescence. This exploratory self-reported survey was developed as an extension of the key informant interviews and was designed to expand upon and quantify each of these factors for intervention development.

Participants

Eligibility criteria included: age 18 - 25 years, TGD identity (self-identifying as TGD or having a gender identity different from sex assigned at birth), and US residency. The sample was

XSL•FO
RenderX

stratified to enroll approximately 50 Black, Indigenous, and People of Color (BIPOC) people and 50 White young people. A total of 104 individuals were surveyed; additional details on survey methodology can be found elsewhere [21].

Assessments

PHQ-2 Depression Status

The psychometrically validated PHQ-2 assessed the frequency of depressed mood and anhedonia over the past two weeks [22]. The two items were (1) "little interest or pleasure in doing things" and (2) "feeling down, depressed, or hopeless." The response options were 0=not at all, 1=several days, 2=more than half the days, 3=nearly every day. The two items were highly correlated (r=0.81; P<.001) and summed. Scores ranged from 0 - 6, and a score ≥3 was considered a positive screen for depression, based on a previously validated threshold for screening for major depression (sensitivity: 83%, specificity: 92%) [22].

Sociodemographic Variables, Family Context, Mental Health Services, Peer Support, and Mental Health Help-Seeking in Adolescence

Sociodemographic variables included age in years (continuous and categorized: 18 - 20, 21 - 23, 23 - 25), race or ethnicity (ie, BIPOC monoracial, BIPOC multiracial, White), gender identity (ie, transgender men, transgender women, nonbinary), US Census region (ie, Northeast, Midwest, South, West), geographic context (ie, urban, suburban, rural), health insurance status (ie, private, public, both, none), and self-identified religion (eg, Agnostic, Atheist, Baptist, Buddhist, Catholic, Christian, Evangelical Protestant, Hindu, Jehovah's Witness, Jewish, Latter-day Saint or Mormon, Mainline Protestant, Muslim, Orthodox Christian, Pagan, Sikh, Spiritual but not religious, Unitarian/Universalist, Unaffiliated, Other, not applicable; coded as Christian, not Christian, unaffiliated, more than one religion, or not applicable).

Family-related variables included family support, assessed using the question: "How much trouble did your parents or caregivers have accepting your gender identity?" Responses were coded as high family support (none/a little), low family support (a moderate amount/a lot), or not applicable (not applicable/they don't know I'm trans); and family religion was self-reported by the young adult (using the same categories as for self-identified religion and coded as Christian, not Christian, unaffiliated, more than one religion, or not applicable).

Participants were queried on mental health services and peer support, including receiving care from a mental health therapist (yes, no) and whether they had ever participated in a peer support group (yes, no). Recalled frequency of mental health help-seeking in adolescence was assessed using 0-to-4-point Likert scales: (1) "When you were a teenager, how often were you looking for information about finding mental health support or a therapist?" (0=never to 4=very often; categorized as never/occasionally versus sometimes/often/very often), (2) "How useful would it have been to have reliable information on finding mental health support or a therapist?" (0=not at all to 4=extremely; categorized as not at all/slightly vs moderately/very/extremely), and (3) "How interested were you in getting therapy related to your gender? (0=not at all to 4=extremely; categorized as not at all/slightly, moderately/very/extremely, or not applicable).

Study Size

This exploratory survey was designed to be complementary to key informant interviews conducted with TGD young adults, caregivers, and health care providers as part of the NIH-funded TransHealthGUIDE project. After an initial round of key informant interviews, the research team identified the need for feedback from a larger group of diverse TGD young adults, than those were accessible for the interviews, and designed this survey to capture additional feedback.

Data Analysis

Descriptive statistics (ie, frequencies, percentages, means, standard deviations) summarized variables for the overall sample and by PHQ-2 depression status. Bivariate tests (2-tailed t tests for continuous variables; χ^2 tests for binary or categorical variables) were used to compare participants with and without depression detected using PHQ-2. Poisson regression models with robust variance were fit to estimate prevalence ratios (PR) and 95% CI for correlates of depression [23]. Bivariate models were fit for each variable, followed by a single multivariable model with all variables to estimate adjusted prevalence ratios (aPR). Variables for modeling were selected *a priori* through a review of existing TGD research, conceptual salience, and findings from key informant interviews. Analyses were conducted using R software (version 4.4.3; R Foundation for Statistical Computing) [24] with statistical significance at P<.05 [24].

Ethical Considerations

After review, the study protocol was deemed exempt by the Boston Children's Hospital Institutional Review Board (IRB-P00043127). This study was performed in accordance with the ethical standards outlined in the 1964 Declaration of Helsinki and its subsequent amendments. Participants completed an informed consent process. This was an anonymized survey, and participants indicated consent by opting into the survey. Participants were compensated $3 for completing the survey in accordance with the standards of the Prolific survey platform.

Results

Study Sample

Sample characteristics are displayed in Table 1 for the total sample and stratified by PHQ-2 depression scores. Overall, 44 of 104 (42%) of the sample had clinically elevated depressive symptoms.

XSL•FO
RenderX

Table . Characteristics of a US nationwide sample of Transgender, Nonbinary, and Gender Diverse young adults aged 18 - 25 years, stratified by Patient Health Questionnaire-2 (PHQ-2)[a] depression[b] .

Variables	Participants (N=104)	Depression[b] (n=44), n (%)	No depression[b] (n=60), n (%)	Bivariate test: *t* test (*df*) or χ^2 test (*df*)[c]	*P* value
Age (years), mean (SD)	21.8 (1.9)	22.1 (1.9)	21.6 (1.9)	–1.21 (90.2)[d]	.23
Age group(years), n (%)				2.09 (2)[e]	.35
18 - 20	28 (26.9)	10 (22.7)	18 (30)		
21 - 22	37 (35.6)	14 (31.8)	23 (38.3)		
23 - 25	39 (37.5)	20 (45.5)	19 (31.7)		
Race/ethnicity, n (%)				1.76 (2)[e]	.41
White	55 (52.9)	20 (45.5)	35 (58.3)		
BIPOC[f] monoracial	23 (22.1)	10 (22.7)	13 (21.7)		
BIPOC multiracial	25 (24)	13 (29.5)	12 (20)		
Missing	1 (1)	1 (2.3)	0 (0)		
Gender identity, n (%)				1.60 (2)[e]	.46
Nonbinary	69 (66.3)	29 (65.9)	40 (66.7)		
Transgender man	24 (23.1)	12 (27.3)	12 (20)		
Transgender woman	11 (10.6)	3 (6.8)	8 (13.3)		
US Census region, n (%)				7.69 (3)[e]	.05
Northeast	27 (26.0)	7 (15.9)	20 (33.3)		
Midwest	20 (19.2)	6 (13.6)	14 (23.3)		
South	34 (32.7)	19 (43.2)	15 (25)		
West	23 (22.1)	12 (27.3)	11 (18.3)		
Geography, n (%)				4.99 (3)[e]	.14
Urban	27 (26.0)	13 (29.5)	14 (23.3)		
Suburban	64 (61.5)	29 (65.9)	35 (58.3)		
Rural	10 (9.6)	1 (2.3)	9 (15.0)		
Unknown	3 (2.9)	1 (2.3)	2 (3.3)		
Health insurance, n (%)				6.86 (3)[e]	.08
Private	48 (46.1)	17 (38.6)	31 (51.7)		
Public	40 (38.5)	20 (45.5)	20 (33.3)		
Private and public	4 (3.8)	0 (0)	4 (6.7)		
No insurance	11 (10.6)	7 (15.9)	4 (6.7)		
Missing	1 (1)	0 (0)	1 (1.7)		
Own religion, n (%)[g]				1.75 (4)[e]	.81
Christian	7 (6.7)	2 (4.5)	5 (8.3)		
Not Christian	13 (12.5)	7 (15.9)	6 (10)		
Unaffiliated	68 (65.4)	29 (65.9)	39 (65)		
More than one	4 (3.8)	1 (2.3)	3 (5)		
Not applicable	12 (11.5)	5 (11.4)	7 (11.7)		

XSL•FO
RenderX

Variables	Participants (N=104)	Depression[b] (n=44), n (%)	No depression[b] (n=60), n (%)	Bivariate test: *t* test (*df*) or χ^2 test (*df*)[c]	*P* value
Family support, n (%)				4.18 (2)[e]	.12
High family support	24 (23.1)	6 (13.6)	18 (30)		
Low family support	31 (9.8)	16 (36.4)	15 (25)		
Not applicable/ They don't know I'm trans	49 (47.1)	22 (50)	27 (45)		
Family religion, n (%)[h]				7.35 (4)[e]	.11
Christian	55 (52.9)	30 (68.2)	25 (41.7)		
Not Christian	3 (2.9)	1 (2.3)	2 (3.3)		
Unaffiliated	19 (18.3)	6 (13.6)	13 (21.7)		
More than one religion	16 (15.4)	4 (9.1)	12 (20)		
Not applicable	11 (10.6)	3 (6.8)	8 (13.3)		
Mental health therapist, n (%)				1.25 (1)[e]	.26
No	56 (53.8)	27 (61.4)	29 (48.3)		
Yes	48 (46.2)	17 (38.6)	31 (51.7)		
Peer support group, n (%)				0.55 (1)[e]	.46
No	76 (73.1)	30 (68.2)	46 (76.7)		
Yes	28 (26.9)	14 (31.8)	14 (23.3)		
How often sought mental health/ therapy information in adults aged <18 years, n (%)				1.03 (1)[e]	.31
Never or occasionally	45 (43.3)	16 (36.4)	29 (48.3)		
Sometimes, often, or very often	59 (56.7)	28 (53.6)	31 (51.7)		
How useful would it have been to have reliable information on mental health/ therapy in adults aged <18 years, n (%)				0.49 (1)[e]	.49
Not at all or slightly	16 (15.4)	5 (11.4)	11 (18.3)		
Moderately, very, or extremely	88 (84.6)	39 (88.6)	49 (81.7)		
How interested you were in getting therapy related to gender (age <18 years), n (%)				1.78 (2)[e]	.40
Not at all or slightly	43 (41.3)	13 (29.5)	30 (50)		
Moderately, very, or extremely	53 (51)	28 (63.6)	25 (41.7)		

XSL•FO
RenderX

Variables	Participants (N=104)	Depression[b] (n=44), n (%)	No depression[b] (n=60), n (%)	Bivariate test: *t* test (*df*) or χ^2 test (*df*)[c]	*P* value
Not applicable	8 (7.7)	3 (6.8)	5 (8.3)		

[a]PHQ-2: Two-item Patient Health Questionnaire (score ≥3 is the validated cut-off and indicates clinically elevated depressive symptoms).

[b]Depression was assessed using the PHQ-2.

[c]Fisher's exact tests were used to obtain *P* values for cell sizes ≤5.

[d]*t* test (*df*).

[e]χ^2test (*df*).

[f]BIPOC: Black, Indigenous, and Other People of Color.

[g]Own Religion: Christian (57.1% Catholic, 14.3% Baptist, 14.3% Orthodox, 14.3% Christian); Not Christian (38.5% Pagan, 30.8% Jewish, 23.1% Buddhist, 7.7% Misotheist); Unaffiliated (41.9% Agnostic, 29.1% Atheist, 19.8% Spiritual, 9.3% Unaffiliated). Percentages may not sum to 100% due to rounding.

[h]Family Religion (self-reported by TGD young adults): Christian (49.2% Catholic, 20.6% Orthodox, 11.1% Mainline Protestant, 4.8% Evangelist, 3.2% Latter-day Saints, 3.2% Christian, 3.2% Baptist, 1.6% Muslim, 1.6% Jehovah's Witness, 1.6% Unitarian); Not Christian (66.7% Buddhist, 33.3% Jewish), Unaffiliated (39.3% Atheist, 28.6% Agnostic, 17.9% Spiritual, 14.3% Unaffiliated). Percentages may not sum to 100% due to rounding.

Participants' mean age was 21.8 (SD 1.9) years; a total of 25/104 (24%) were BIPOC multiracial individuals, while 23/104 (22%) were BIPOC monoracial individuals, and 69/104 (66%) participants identified as nonbinary. The majority (n=64, 61%) of participants resided in a suburban area, and 11/104 (11%) did not have health insurance. Regarding their religious affiliation, 68 (65%) participants described themselves as unaffiliated, while 55 (53%) described their family's religion as Christian. Half (49/104, 47%) of the participants were not out to their family, 31 (30%) reported low family support, while 24 (23%) participants reported high family support.

Among the participants, 56 (46%) had received mental health therapy and 28 of 104 (27%) had participated in a peer support group. When asked to recall their mental health needs during adolescence: 59 (57%) participants reported seeking information about mental health support or therapy, 88 (85%) indicated that reliable information on mental health support or therapy would have been useful, and 53 (51%) participants expressed having had an interest in therapy related to their gender.

Correlates of PHQ-2 Depression Score

In bivariate models (Table 2), increased PHQ-2 depression prevalence was associated with (1) being BIPOC multiracial versus White (PR1.39; 95% CI 1.05 - 1.84), (2) having low family support versus high (PR 1.53; 95% CI 1.09 - 2.15), (3) seeking information about mental health therapy in adolescence (PR 1.31; 95% CI 1.02 - 1.67), and (4) being interested in mental health therapy related to gender during adolescence (PR 1.41; 95% CI 1.09 - 2.05). Conversely, residing in a rural versus suburban area was associated with decreased PHQ-2 depression prevalence (PR 0.54; 95% CI 0.32 - 0.92). No other variables reached statistical significance in bivariate models.

XSL•FO
RenderX

Table . Bivariate and multivariable Poisson regression models with robust variance estimation: Correlates of Patient Health Questionnaire-2 (PHQ-2)[a] depression scores in a nationwide sample of transgender, nonbinary, and gender diverse young adults aged 18 - 25 across the United States.

Variables	Bivariate models (N=103)[b], PR[c] (95% CI)	*P* value	Multivariable model (N=103)[b], aPR[d] (95% CI)	*P* value
Age groups (years)				
18 - 20	Ref		Ref	
21 - 22	0.99 (0.72-1.35)	.94	0.83 (0.55-1.24)	.35
23 - 25	1.20 (0.89-1.61)	.24	1.03 (0.69-1.56)	.87
Race/Ethnicity				
White	Ref		Ref	
BIPOC[e] monoracial	1.10 (0.81-1.50)	.54	0.96 (0.63-1.45)	.84
BIPOC multiracial	1.39 (1.05-1.84)	.03	1.04 (0.75-1.44)	.81
Gender identity				
Nonbinary	Ref		Ref	
Transgender man	1.06 (0.80-1.40)	.70	0.91 (0.64-1.29)	.59
Transgender woman	0.71 (0.45-1.12)	.14	0.69 (0.41-1.16)	.16
US Census region				
Northeast	Ref		Ref	
Midwest	1.19 (0.82-1.73)	.36	1.16 (0.75-1.80)	.51
South	1.31 (0.94-1.80)	.11	1.02 (0.69-1.50)	.94
West	1.31 (0.92-1.87)	.13	1.14 (0.75-1.71)	.54
Geography				
Suburban	Ref		Ref	
Urban	0.96 (0.73-1.27)	.79	0.98 (0.72-1.36)	.93
Rural	0.54 (0.32-0.92)	.02	0.48 (0.26-0.92)	.03
Unknown	1.21 (0.64-2.28)	.57	1.16 (0.58-2.35)	.67
Health insurance				
Private	Ref		Ref	
Public	1.28 (0.99-1.65)	.07	1.08 (0.80-1.46)	.63
Private and public	0.86 (0.42-1.76)	.67	0.91 (0.40-2.05)	.82
No Insurance	1.36 (0.93-1.99)	.11	1.01 (0.63-1.62)	.97
Own religion				
Unaffiliated	Ref		Ref	
Not Christian	1.26 (0.90-1.76)	.19	1.66 (1.04-2.63)	.03
Christian	0.78 (0.45-1.34)	.36	0.76 (0.41-1.41)	.39
More than one religion	0.97 (0.51-1.84)	.93	0.78 (0.38-1.60)	.39
Not applicable	1.07 (0.74-1.55)	.73	1.25 (0.80-1.93)	.32
Family support				
High family support	Ref		Ref	
Low family support	1.53 (1.09-2.15)	.01	1.54 (1.05-2.27)	.03
Not applicable/ they don't know I'm trans	1.17 (0.85-1.62)	.34	1.06 (0.69-1.60)	.80
Family religion				
Unaffiliated	Ref		Ref	

XSL•FO
RenderX

Variables	Bivariate models (N=103)[b], PR[c] (95% CI)	*P* value	Multivariable model (N=103)[b], aPR[d] (95% CI)	*P* value
Not Christian	0.75 (0.32-1.74)	.50	1.61 (0.22-1.68)	.34
Christian	1.09 (0.80-1.49)	.59	1.05 (0.70-1.58)	.80
More than one religion	0.88 (0.58-1.35)	.57	0.86 (0.49-1.51)	.60
Not applicable	0.61 (0.36-1.04)	.07	0.65 (0.33-1.25)	.20
Mental health therapist				
No	Ref		Ref	
Yes	0.92 (0.73-1.17)	.51	0.71 (0.53-0.97)	.03
Peer support group				
No	Ref		Ref	
Yes	1.20 (0.93-1.55)	.16	0.97 (0.68-1.37)	.86
How often sought information on mental health/ therapy by individuals aged<18 years				
Never or occasionally	Ref		Ref	
Sometimes, often, or very often	1.31 (1.02-1.67)	.03	1.15 (0.79-1.66)	.46
How useful would It would have been to have reliable information on mental health/ therapy in individuals aged <18 years				
Not at all or slightly	Ref		Ref	
Moderately, very, or extremely	1.33 (0.92-1.91)	.13	1.18 (0.73-1.92)	.50
How interested you were in getting therapy related to gender in adults aged<18 years				
Not at all or slightly	Ref		Ref	
Moderately, very, or extremely	1.41 (1.09-1.82)	.01	1.47 (0.98-2.21)	.06
Not applicable	1.29 (0.81-2.05)	.29	1.33 (0.80-2.23)	.28

[a]PHQ-2 score ≥3 indicating clinically elevated depressive symptoms (validated cut-off).

[b]Models include N=103 participants; 1 participant was excluded due to missing data on health insurance.

[c]PR: prevalence ratio.

[d]aPR: adjusted prevalence ratio.

[e]BIPOC: Black, Indigenous, and Other People of Color.

In a multivariable model (Table 2), after adjusting for all other variables, identifying one's own religion as not Christian versus unaffiliated (aPR 1.66; 95% CI 1.04 - 2.63) and low family support (aPR 1.54; 95% CI 1.05 - 2.27) were associated with increased prevalence of positive depression screening. Residing in a rural versus suburban area (aPR 0.48; 95% CI 0.26 - 0.92) and ever receiving mental health therapy (aPR 0.71; 95% CI 0.53 - 0.97) were associated with decreased PHQ-2 depression prevalence in this model. Age, race or ethnicity, gender identity, US Census region, health insurance, family religion, peer support, history of interest in mental health therapy for gender during adolescence, frequency of information-seeking about mental health resources in adolescence, and desire for reliable mental health information during adolescence, were not significantly correlated with depression based on PHQ-2 scores.

Discussion

Principal Results

In this nationwide online study, depression identified using PHQ-2 scores was highly prevalent, with more than 4 in 10 TDG young adults showing clinically elevated symptoms. Low levels of family support and nonChristian religious affiliation were associated with an increased likelihood of positive PHQ-2 depression screening. This exploratory study highlights the high rates of depressive symptoms among TGD young adults as well as the need for tailored interventions to improve mental health among those at the highest risk.

Comparison With Prior Work

This study found that TGD young people with low family support had a higher prevalence of PHQ-2 depression compared to those with highly supportive families. This finding corroborates prior research on the protective role of family support on TGD young peoples' mental health and has informed intervention development for the broader TransHealthGUIDE

XSL•FO
RenderX

study [15,16]. For example, a key intervention component that has been developed for the overall study includes app-based education for TGD young adults and their caregivers on ways to improve family communication, connection, and support.

No differences in PHQ-2 depression prevalence were found for family religion, despite religious affiliation influencing attitudes toward TGD issues in the United States [18,19]. The TGD young people identifying their own religion as nonChristian (eg, Jewish, Muslim, Buddhist) had a higher prevalence of PHQ-2 depression than those identifying as unaffiliated; however, no significant differences were found between Christian and unaffiliated TGD young adults. The association between religious affiliation and depression has been inconsistent in prior research on young people [25], including a recent systematic review of religion, spirituality, and mental health among TGD youth [17]. Future mental health research may benefit from measuring not only religious affiliation but also the degree of religiosity, which was not assessed in this study.

This study found that a history of mental health help-seeking, specifically in young adults receiving mental health therapy was significantly associated with reduced PHQ-2 depression prevalence. This highlights the need for interventions that connect TGD young adults to necessary mental health services. The American Academy of Pediatrics recommends that adolescents aged 12 years and older be screened annually for depression using a formal self-report screening tool to identify and treat those with depression early in the course of illness, potentially reducing long-term depression-related morbidities [26]. This finding underscores the importance of improving mental health care access for TGD young adults to address depression, including telehealth and other online or digitally delivered services and interventions [27].

Regarding sociodemographic differences, we found no evidence of gender identity differences in PHQ-2 depression prevalence, inconsistent with prior research demonstrating the highest depression burden among nonbinary young adults [9]. However, previous studies have reported mixed findings regarding gender identity differences in depression [13], suggesting the need for ongoing research. While BIPOC multiracial TGD young people exhibited elevated PHQ-2 depression in bivariate analyses, consistent with national US population data, this finding did not remain significant in the adjusted model [3]. Geographic context emerged as a significant factor in this study, in alignment with psychiatric epidemiologic research showing lower depression rates in rural settings in high-income countries [28]. The TGD young adults residing in rural areas had lower PHQ-2 depression prevalence compared to those living in suburban areas. Prior studies have found greater mental distress symptoms among rural versus nonrural TGD adults but did not distinguish between the urban and suburban subgroups [29]. More recent research identified the lowest levels of depressive symptoms in rural versus urban-dwelling TGD adults [30]. Additional research is needed to better understand PHQ-2 depression prevalence and set of symptoms among TGD young adults in rural, urban, and suburban settings.

Limitations

These findings should be interpreted alongside this study's limitations. First, analyses were exploratory, and the small sample size limits statistical power to detect significant associations; the findings will require replication. Further, given the cross-sectional design, the study findings are associational only, and longitudinal research is recommended. Second, related to study measures, the global measure of family support may have overlooked important details about family dynamics, such as communication, acceptance, and explicit expressions of care and support [31]. Relatedly, mental health care utilization variables were self-reported rather than objectively measured. Future research that includes self-reported data alongside electronic health record data would strengthen the evidence base. Third, this study did not assess nonpsychiatric aspects of participants' medical histories, and unmeasured confounding due to gender-affirming medical treatments cannot be ruled out [10,12]. Lastly, there are few validated measures for assessing mental health help-seeking behaviors [32]. Further research to understand online help-seeking behaviors for TGD young people using validated measures is warranted to inform future interventions [33].

Conclusions

The high prevalence of PHQ-2 depression in this online sample of TGD young adults underscores the importance of universal screening for depression in this population. It also highlights the need for appropriate systems to ensure prompt diagnosis, treatment, and follow-up for TGD young adults with PHQ-2 depression. Further, tailored interventions may be necessary for subgroups at increased risk for depression, such as TGD young adults with low family support, to decrease mental health disparities. The continued expansion of online interventions may facilitate the engagement of TGD young adults who would benefit from mental health prevention and treatment efforts [27]. This study's findings have important implications for clinical care and the design, development, and implementation of interventions aimed at improving mental health outcomes and reducing health disparities among TGD young adults.

Acknowledgments

The authors wish to thank Jessica Kremen, MD, Meg Quint, BS, and the Community Advisory Board for their feedback on the survey questionnaire.

Research reported in this publication was supported by the National Institutes of Health (NIH) from the National Institute of Mental Health under award number U01MH136558 (MPIs: RX & SR) and the National Institute on Minority Health and Health Disparities of the NIH (R01MD015031; PI: SLK-W). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

XSL•FO
RenderX

Data Availability

Requests for the dataset and code for this study can be submitted to the corresponding author for consideration.

Authors' Contributions

Conceptualization: ERB, SR, RX
Data curation: ERB, SR, YL
Formal analysis: SR, YL
Funding acquisition: RX
Investigation: ERB, KK, RT, RX, SR
Project administration: KK, RT, RX, SR
Resources: RX, SR
Supervision: RX, SR
Validation: SR, YL,
Visualization: SR, YL
Writing – original draft: RX, SR, YL
Writing – review & editing: ASK, ERB, KK, MRG, RT, RX, SLK-W, SR, SWC, YL

Conflicts of Interest

SLR and ASK receive royalties from McGraw Hill for co-editing the textbook, “Transgender and Gender Diverse Health Care: The Fenway Guide.” ASK receives royalties from American Psychiatric Association Publishing for co-editing the textbook, “Gender-Affirming Psychiatric Care.” SWC consults for or have consulted in the recent past for several universities, nonprofit organizations, government organizations, and Viiv Healthcare, none of which have any involvement in this manuscript. SLK-W is a consultant for Paramount Global, who was not involved in the work reported in this manuscript.

References

1. Lee B, Wang Y, Carlson SA, et al. National, state-level, and county-level prevalence estimates of adults aged ≥18 years self-reporting a lifetime diagnosis of depression - United States, 2020. MMWR Morb Mortal Wkly Rep 2023 Jun 16;72(24):644-650. [doi: 10.15585/mmwr.mm7224a1] [Medline: 37318995]
2. Zhang Z, Jackson SL, Gillespie C, Merritt R, Yang Q. Depressive symptoms and mortality among US adults. JAMA Netw Open 2023 Oct 2;6(10):e2337011. [doi: 10.1001/jamanetworkopen.2023.37011] [Medline: 37812418]
3. Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the united states: results from the 2021 national survey on drug use and health (HHS publication nopep22-07-01-005, NSDUH series H-57). : Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration; 2022 URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt39443/2021NSDUHFFRRev010323.pdf [accessed 2025-03-21]
4. Solmi M, Radua J, Olivola M, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry 2022 Jan;27(1):281-295. [doi: 10.1038/s41380-021-01161-7] [Medline: 34079068]
5. Siu AL, US Preventive Services Task Force (USPSTF), Bibbins-Domingo K, et al. Screening for depression in adults: US Preventive Services Task Force Recommendation Statement. JAMA 2016 Jan 26;315(4):380-387. [doi: 10.1001/jama.2015.18392] [Medline: 26813211]
6. US Preventive Services Task Force, Barry MJ, Nicholson WK, et al. Screening for depression and suicide risk in adults: US Preventive Services Task Force Recommendation Statement. JAMA 2023 Jun 20;329(23):2057-2067. [doi: 10.1001/jama.2023.9297] [Medline: 37338872]
7. Reynolds CF 3rd, Frank E. US Preventive Services Task Force Recommendation Statement on screening for depression in adults: not good enough. JAMA Psychiatry 2016 Mar;73(3):189-190. [doi: 10.1001/jamapsychiatry.2015.3281] [Medline: 26815331]
8. Reisner SL, Vetters R, Leclerc M, et al. Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. J Adolesc Health 2015 Mar;56(3):274-279. [doi: 10.1016/j.jadohealth.2014.10.264] [Medline: 25577670]
9. Lipson SK, Raifman J, Abelson S, Reisner SL. Gender minority mental health in the U.S.: results of a national survey on college campuses. Am J Prev Med 2019 Sep;57(3):293-301. [doi: 10.1016/j.amepre.2019.04.025] [Medline: 31427032]
10. Tordoff DM, Wanta JW, Collin A, Stepney C, Inwards-Breland DJ, Ahrens K. Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. JAMA Netw Open 2022 Feb 1;5(2):e220978. [doi: 10.1001/jamanetworkopen.2022.0978] [Medline: 35212746]
11. Eisenberg ME, Gower AL, McMorris BJ, Rider GN, Shea G, Coleman E. Risk and protective factors in the lives of transgender/gender nonconforming adolescents. J Adolesc Health 2017 Oct;61(4):521-526. [doi: 10.1016/j.jadohealth.2017.04.014] [Medline: 28736148]

XSL•FO
RenderX

12. Green AE, DeChants JP, Price MN, Davis CK. Association of gender-affirming hormone therapy with depression, thoughts of suicide, and attempted suicide among transgender and nonbinary youth. J Adolesc Health 2022 Apr;70(4):643-649. [doi: 10.1016/j.jadohealth.2021.10.036] [Medline: 34920935]
13. Wittlin NM, Kuper LE, Olson KR. Mental health of transgender and gender diverse youth. Annu Rev Clin Psychol 2023 May 9;19:207-232. [doi: 10.1146/annurev-clinpsy-072220-020326] [Medline: 36608332]
14. Dhanani LY, Totton RR. Have you heard the news? The effects of exposure to news about recent transgender legislation on transgender youth and young adults. Sex Res Social Policy 2023 Apr 5;2023:1-15. [doi: 10.1007/s13178-023-00810-6] [Medline: 37363351]
15. Simons L, Schrager SM, Clark LF, Belzer M, Olson J. Parental support and mental health among transgender adolescents. J Adolesc Health 2013 Dec;53(6):791-793. [doi: 10.1016/j.jadohealth.2013.07.019] [Medline: 24012067]
16. Pariseau EM, Chevalier L, Long KA, Clapham R, Edwards-Leeper L, Tishelman AC. The relationship between family acceptance-rejection and transgender youth psychosocial functioning. Clin Pract Pediatr Psychol 2019 Sep;7(3):267-277. [doi: 10.1037/cpp0000291]
17. Lekwauwa R, Funaro MC, Doolittle B. Systematic review: The relationship between religion, spirituality and mental health in adolescents who identify as transgender. J Gay Lesbian Ment Health 2023 Oct 2;27(4):421-438. [doi: 10.1080/19359705.2022.2107592]
18. Attitudes about transgender issues vary widely among Christians, religious 'nones' in US. Pew Research Center. 2022. URL: https://www.pewresearch.org/short-reads/2022/07/07/attitudes-about-transgender-issues-vary-widely-among-christians-religious-nones-in-u-s/
19. Campbell M, Hinton JDX, Anderson JR. A systematic review of the relationship between religion and attitudes toward transgender and gender-variant people. Int J Transgend 2019;20(1):21-38. [doi: 10.1080/15532739.2018.1545149] [Medline: 32999592]
20. Uhlhaas PJ, Davey CG, Mehta UM, et al. Towards a youth mental health paradigm: a perspective and roadmap. Mol Psychiatry 2023 Aug;28(8):3171-3181. [doi: 10.1038/s41380-023-02202-z] [Medline: 37580524]
21. Boskey ER, Quint M, Xu R, et al. Gender affirmation-related information-seeking behaviors in a diverse sample of transgender and gender-diverse young adults: survey study. JMIR Form Res 2023 Aug 15;7:e45952. [doi: 10.2196/45952] [Medline: 37581925]
22. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
23. Petersen MR, Deddens JA. A comparison of two methods for estimating prevalence ratios. BMC Med Res Methodol 2008 Feb 28;8:9. [doi: 10.1186/1471-2288-8-9] [Medline: 18307814]
24. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2024. URL: https://www.R-project.org
25. Aggarwal S, Wright J, Morgan A, Patton G, Reavley N. Religiosity and spirituality in the prevention and management of depression and anxiety in young people: a systematic review and meta-analysis. BMC Psychiatry 2023 Oct 10;23(1):729. [doi: 10.1186/s12888-023-05091-2] [Medline: 37817143]
26. Zuckerbrot RA, Cheung A, Jensen PS, Stein REK, Laraque D, GLAD-PC STEERING GROUP. Guidelines for adolescent depression in primary care (GLAD-PC): Part I. Practice preparation, identification, assessment, and initial management. Pediatrics 2018 Mar;141(3):e20174081. [doi: 10.1542/peds.2017-4081] [Medline: 29483200]
27. Hamnvik OPR, Agarwal S, AhnAllen CG, Goldman AL, Reisner SL. Telemedicine and inequities in health care access: the example of transgender health. Transgend Health 2022 Apr;7(2):113-116. [doi: 10.1089/trgh.2020.0122] [Medline: 36644516]
28. Xu C, Miao L, Turner D, DeRubeis R. Urbanicity and depression: a global meta-analysis. J Affect Disord 2023 Nov 1;340:299-311. [doi: 10.1016/j.jad.2023.08.030] [Medline: 37557989]
29. Horvath KJ, Iantaffi A, Swinburne-Romine R, Bockting W. A comparison of mental health, substance use, and sexual risk behaviors between rural and non-rural transgender persons. J Homosex 2014;61(8):1117-1130. [doi: 10.1080/00918369.2014.872502] [Medline: 24380580]
30. Ralston AL, Holt NR, Andrews AR 3rd, et al. Mental health and marginalization stress in transgender and gender diverse adults: differences between urban and non-urban experiences. Psychol Sex Orientat Gend Divers 2024 Mar;11(1):165-176. [doi: 10.1037/sgd0000595] [Medline: 38577413]
31. McGregor K, Rana V, McKenna JL, Williams CR, Vu A, Boskey ER. Understanding family support for transgender youth: impact of support on psychosocial functioning. J Adolesc Health 2024 Aug;75(2):261-266. [doi: 10.1016/j.jadohealth.2024.04.006] [Medline: 38842989]
32. Rickwood D, Thomas K. Conceptual measurement framework for help-seeking for mental health problems. Psychol Res Behav Manag 2012;5:173-183. [doi: 10.2147/PRBM.S38707] [Medline: 23248576]
33. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873. [doi: 10.2196/13873] [Medline: 31742562]

XSL•FO
RenderX

Abbreviations

aPR: adjusted prevalence ratio
BIPOC: Black, Indigenous, and People of Color
PHQ-2: two-item patient health questionnaire
PR: prevalence ratio
TGD: Transgender, nonbinary, gender diverse

Edited by TDA Cardoso; submitted 18.09.24; peer-reviewed by K Clark, ZJG Regencia; revised version received 28.01.25; accepted 24.02.25; published 24.03.25.

Please cite as:
Reisner S, Liu Y, Tham R, Kane K, Cole SW, Boskey ER, Katz-Wise SL, Keuroghlian AS, Xu R
Prevalence and Correlates of Clinically Elevated Depressive Symptoms in a Nationwide Sample of Transgender, Nonbinary, and Gender Diverse Young Adults in the United States: Cross-Sectional Survey Study
Interact J Med Res 2025;14:e66630
URL: https://www.i-jmr.org/2025/1/e66630
doi:10.2196/66630

© Sari Reisner, Yuxin Liu, Regina Tham, Kaiden Kane, S Wilson Cole, Elizabeth R Boskey, Sabra L Katz-Wise, Alex S Keuroghlian, Rena Xu. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 24.3.2025. 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

Citizen Worry and Adherence in Response to Government Restrictions in Switzerland During the COVID-19 Pandemic: Repeated Cross-Sectional Online Surveys

Vanessa Kraege[1,2*], MD, MBA; Céline Dumans-Louis[2,3*], MD; Céline Maglieri[1], MSc; Séverine Bochatay[4]; Marie-Anne Durand[2,5,6], Prof Dr; Antoine Garnier[7,8], MBA, MD; Kevin Selby[2,5], MD; Christian von Plessen[2,5], MD, PhD

[1]Innovation and Clinical Research Directorate, University Hospital of Lausanne, Lausanne, Switzerland

[2]Faculty of Biology and Medicine, University of Lausanne, Lausanne, Switzerland

[3]Department of Internal Medicine, Le Groupement Hospitalier de l'Ouest Lémanique, 1260 Nyon, Switzerland

[4]Faculty of Arts, University of Lausanne, Lausanne, Switzerland

[5]Department of Ambulatory Care, Centre universitaire de médecine générale et santé publique, Lausanne, Lausanne, Switzerland

[6]The Dartmouth Institute for Health Policy and Clinical Practice, Dartmouth College, Hanover, NH, United States

[7]Faculty of Science and Medicine, University of Fribourg, Fribourg, Switzerland

[8]Department of Medicine and Specialities, Fribourg Cantonal Hospital, Fribourg, Switzerland

[*]these authors contributed equally

Corresponding Author:
Céline Dumans-Louis, MD
Department of Internal Medicine
Le Groupement Hospitalier de l'Ouest Lémanique
Chemin du Monastier 10
1260 Nyon
Switzerland
Phone: 41 765567810
Email: celine.dumanslouis@ghol.ch

Abstract

Background: Good communication between health authorities and citizens is crucial for adherence to preventive measures during a pandemic. Crisis communication often appeals to worries about negative consequences for oneself or others. While worry can motivate protective behavior, it can also be overwhelming and lead to irrational choices or become a mental health problem. Also, the levels and consequences of worry can differ between different groups of citizens. Little is known about the evolution of worries during the pandemic and adherence to measures in distinct groups.

Objective: This study aimed to evaluate worries in the Swiss population as well as associations between worry levels and citizens' adherence to government restrictions during different phases of the COVID-19 pandemic.

Methods: We carried out an observational study with 4 cross-sectional online surveys of adults in the Canton of Vaud, Switzerland. Questionnaires were distributed through social media and websites during 4 periods: survey 1: April 17 to May 14, 2020; survey 2: May 15 to June 22, 2020; survey 3: October 30 to December 12, 2020; and survey 4: June 18 to December 30, 2021. On visual analog scales from 0 to 100, participants reported worry, self-adherence to pandemic restrictions, and their perceived adherence to others. We used multivariable linear regression, adjusting for age, gender, health literacy, and education to assess associations between self-reported worry, adherence, and study periods.

Results: We collected 7106 responses. After excluding 2377 questionnaires (incomplete, age <18 years, residence outside Vaud), 4729 (66.55%) were analyzed (mean age 47, SD 15.6 years, 63.96% women). Mean worry across the 4 periods was 42/100, significantly higher in women (44.25/100, vs 37.98/100; P<.001) and young people (43.77/100 in those aged 18-39 years, vs 41.69/100; P=.005; in those aged 40-64 years and 39.16/100; P=.002; in those aged >64 years). Worries were higher during survey 1 and survey 3 (52.41/100 and 56.32/100 vs 38.93/100, P<.001; and 35.71/100, P<.001) than during survey 2 and survey 4, respectively. This corresponds to pandemic peaks during which federal restrictions were better followed with self-reported adherence of 84.80/100 and 89.59/100 in survey 1 and survey 3 versus 78.69/100 (P<.001) and 78.64/100 (P<.001) in survey 2

XSL•FO
RenderX

and survey 4. A 2.9-point increase in worry score, adjusted for the pandemic period, gender, age, education, and health literacy, was associated with a 10-point increase in personal adherence score (95% CI 2.5-3.2; *P*<.001).

Conclusions: Worries were higher in women, young people, and during the peak of the COVID-19 pandemic. Higher worry levels were associated with increased self-reported adherence to federal restrictions. Authorities should consider population worry levels and population subgroups in the planning and design of pandemic communication.

(Interact J Med Res 2025;14:e55636) doi:10.2196/55636

KEYWORDS

COVID-19 pandemic; citizens; worry; anxiety; communication; prevention; adherence; restrictions; Switzerland; cross sectional; online survey; survey; Swiss; adults; questionnaire; social media; linear regression; age; gender; health literacy; education; women; young people

Introduction

Effective communication between health authorities and the population is crucial to achieving public health goals during a pandemic. Providing clear, consistent, and reliable information that motivated behavior changes without triggering resistance was a major challenge during the COVID-19 pandemic. Sanitary restrictions were often rapidly issued and modified to contain the spread of the disease [1-3]. Citizens were expected to make drastic behavioral changes.

Public health authorities stressed the seriousness and risks of the pandemic to justify restrictions and encourage citizens' adherence to them. In support of such an approach, the Health Belief Model argues that preventive health behaviors are influenced by perceived susceptibility to illness, the severity of the disease, benefits of and barriers to health-promoting actions, cues to action, as well as self-efficacy [4]. Also, during the COVID-19 pandemic, protective behaviors were associated with these factors, especially when the "perceived benefit" of a measure was clear [5].

Appealing to emotions such as fear can hence be a persuasive way of motivating respect for protective measures [6,7]. Indeed, fear leads to behavioral change if people feel capable of dealing with the threat, while they become defensive when feeling helpless and incapable of acting [8-10]. Overdriven or ill-conceived fear-based communication may even provoke counterproductive behavior.

Levels of anxiety, worry, and stress were high during the pandemic. According to a systematic review and meta-analysis, anxiety prevalence was around 30% worldwide after the first COVID-19 wave [11]. Others have confirmed these findings [12,13]. Among professionally active persons, 42% of participants reported being worried about the COVID-19 pandemic in August and September 2020 [14]. Young adults in the city of Zurich, Switzerland, reported elevated stress levels in April 2020, in the aftermath of the first wave [15], as was found more generally in Swiss adults too [16]. In terms of risk factors, anxiety was higher in women, younger people, and vulnerable persons [17-19].

In our previously published cross-sectional population survey, performed during the first wave of the pandemic, we found high self-reported adherence to official restrictions, which increased with age and level of worry [20]. As in the aforementioned studies, worry was high, particularly among people in isolation and with lower health literacy. Nearly half of the respondents felt that government responses were adequate or, associated with higher levels of worry, even insufficient. Neither the aforementioned nor our cross-sectional study could determine the evolution of these associations throughout the pandemic.

Thus, we conducted surveys during different phases of the pandemic to describe the evolution of worries in the Swiss population as well as associations between worries and adherence to governmental restrictions. Our overall aim was to contribute new insights to this understudied area to help improve crisis communication during future pandemics.

Methods

Study Design and Setting

We conducted repeated cross-sectional online surveys in Vaud, a French-speaking canton of 823,000 inhabitants (2021) in Switzerland. We launched 4 surveys between April 2020 and December 2021: survey 1 between April 17 and May 14, 2020 (4 weeks); survey 2 between May 15 and June 22, 2020 (5.5 weeks); survey 3 between October 30 and December 1, 2020 (4.5 weeks); and survey 4 between June 18 and December 30, 2021 (28 weeks, Table 1). Some of the survey items were adapted or replaced to capture changes in federal measures. We followed the CHERRIES (Checklist for Reporting Results of Internet e-Surveys) guidelines [21]. Self-reported worry was an outcome of the study of worry levels and an exposure variable for the study of associations between worry and self-reported adherence during these 4 COVID-19 pandemic periods.

XSL•FO
RenderX

Table 1. Surveyed periods and sentinel pandemic-related events.

Survey	Period	Number of weeks	Sentinel events
Survey 1: End of the first pandemic wave	April 17, 2020, to May 14, 2020	4	• March 16, 2020: Semiconfinement, only essential shops open, gatherings of a maximum of 5 people • April 27, 2020: Partial reopening of shops • May 11, 2020: Reopening of schools
Survey 2: After the first pandemic wave	May 15, 2020, to June 22, 2020	5.5	• June 19, 2020: End of an extraordinary situation
Survey 3: During the second pandemic wave	October 30, 2020, to December 1, 2020	4.5	• Mandatory wearing of masks in indoor public spaces; gatherings limited to 15 people
Survey 4: Following pandemic waves	June 18, 2021, to December 30, 2021	28	• Vaccination available to all, use of COVID-19 vaccination certificate

Table 1 presents the COVID-19 waves and their duration in the French-speaking part of Switzerland. The first and second forms were distributed at the end of the first wave, corresponding to the gradual emergence from confinement. The third form was distributed over a longer period, which included the second wave and the resumption of restrictive measures. The fourth form was distributed once vaccination was available for the entire population.

Participant Recruitment

Using a weblink, we distributed the surveys on the social media platforms of multiple community organizations to collect a convenience sample of the population. These organizations were a regional consumer organization, regional disease leagues for cancer and diabetes, the association of senior citizens as well as the cantonal websites for the COVID-19 testing and vaccination decisions. These cantonal sites were used by large portions of the population. The organizations advertised the study through links on their websites and some social media accounts. The links were accompanied by a short explanation of the study and its purpose. No incentives to participate were given. The online interface for the survey was created in REDCap (Research Electronic Data Capture, Vanderbilt University).

Surveys

The development and testing of the survey are described in our previous publication with results from survey 1 [20]. Each survey was submitted to 5 nonmedical persons to test the understandability of questions. There was no review step for this short questionnaire. The first, second, and fourth surveys had 20 items, and the third survey had 25 items. A total of 12 items remained unchanged throughout all surveys. We included demographic data, such as age, sex, number of persons per household, canton of residence, level of education, literacy, and whether the respondent had been tested for the COVID-19 pandemic. For the literacy question, we used a validated item from Chew et al [22]. Employment status was included in surveys 2-4. Respondents rated (1) worry about the pandemic situation, (2) self-reported adherence to government restrictions, and (3) perceived adherence of others to government restrictions, on visual analog scales from 0 to 100 (0=not at all; 100=in all situations). Items were not randomized. Participants could go back to earlier questions at any time. Adaptive questioning was used for several items. The first 2 surveys took up 6 screens, and the 2 latter, 7. REDCap automatically generated a completeness variable if participants went all the way to the end of the survey. Anyone who opened the survey generated a response. We did not determine unique site visitors, establish view or participation rates, or IP address checks, as surveys were entirely anonymous and the risk of repeating them was low. The time used to fill them in was not registered. No cookies were used, and there were no other techniques to analyze the log file of our database. Statistical corrections were not used. The 4 surveys can be found in Multimedia Appendices 1-4.

Statistical Analyses

The item “Level of education” was dichotomized into “university or college education” or other. Health literacy was dichotomized according to ease of answering a medical form on one’s own: low literacy (“never,” “rarely,” or “sometimes” at ease) and “high literacy” (“often” or “always”). This was based on a validated item [22] and our previous article [20].

We limited our analysis to complete questionnaires for 2 reasons. First, the survey was distributed using an online link on government websites and a large number of persons clicked the link but only completed 1 or 2 questions. Second, no single question had many missing responses and we preferred to maintain consistency across surveys. We calculated means with SD and frequencies with IQR as appropriate. Independence between surveys was tested with the *t* test for continuous variables (eg, age), and with the chi-square test for gender, education, and health literacy. We performed linear regressions to analyze associations between the 4 periods and level of worry (model A), self-reported adherence (model B), and perceived adherence of others to restrictions (model C). First, we performed univariate linear regression, followed by multivariable linear regression controlling for age (grouped as 18-39 years, 40-64 years, and 65 years or older), sex (dichotomized male-female), level of education (dichotomized university education or other), and health literacy (dichotomized high or low health literacy). We subsequently used the margins command in Stata (StataCorp) to report absolute differences in the predicted levels of worry or adherence with each model. The level of significance was set to $P<.05$. Statistical analyses were performed in Microsoft Excel and Stata (version 16.1).

XSL•FO
RenderX

Ethical Considerations

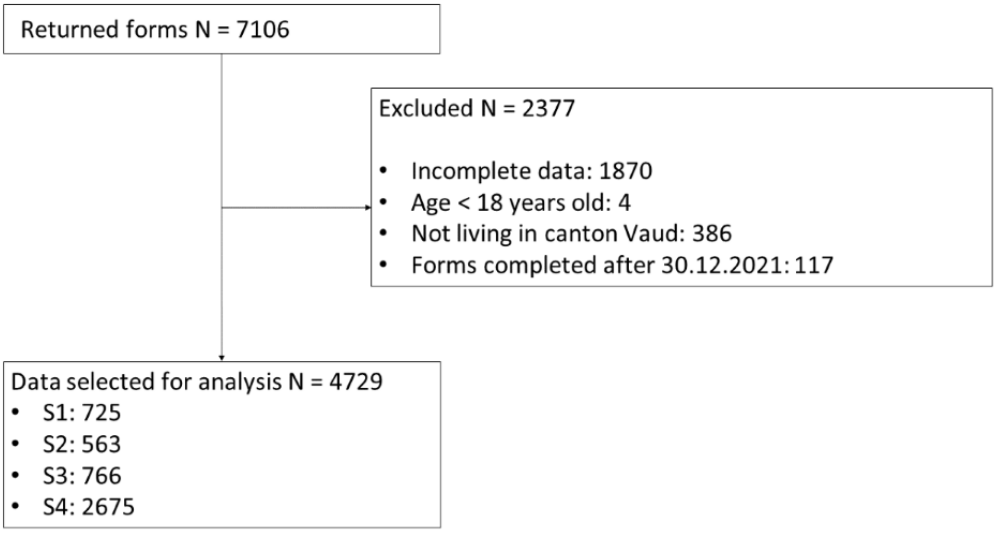
According to the Cantonal Commission on Ethics in Research Involving Human Beings of the Canton of Vaud, Switzerland the study was exempted from ethical review because it did not qualify as human subjects research and all data collection was anonymous (2024-010901).

Results

Participant Characteristics

Citizens completed 7106 surveys between April 17, 2020, and December 20, 2021. After the exclusion of minors, persons living outside the Canton, and incomplete questionnaires, 4729 (66%) surveys remained. The number of questionnaires per period ranged between 563 and 2675 (Figure 1).

Figure 1. Study participants.



The participants were more often women (3025/4729, 63.96%) and between 40 and 64 years old (2442/4729, 51.64%). Furthermore, 2526/4729 (53.42%) of participants had attended university or college, and 4510/4729 (95.37%) reported high health literacy. The participants were younger in survey 3 (mean age 43.2, SD 14.6 years; P<.001) than in the other surveys (survey 1: mean age 47.9, SD 14.6 years; survey 2: mean age 47.3, SD 15.6 years; and survey 4: mean age 48.7, SD 15.3 years; Table 2).

XSL•FO
RenderX

Table 2. Demographics, education, health literacy, and employment of participants (n=4729).

Variables	All surveys	Survey 1[a]	Survey 2[b]	Survey 3[c]	Survey 4[d]
Age (years), n (%)					
18-39	1574 (33.3)	244 (33.7)	202 (35.9)	325 (42.4)	803 (30)
40-64	2442 (51.6)	365 (50.3)	268 (47.6)	369 (48.2)	1440 (53.8)
≥65	713 (15.1)	116 (16)	93 (16.5)	72 (9.4)	432 (16.1)
Years, mean (SD)	47.5 (15.3)	47.9 (14.6)	47.3 (15.6)	43.2 (14.6)	48.7 (15.3)
Gender, n (%)					
Male	1698 (35.9)	168 (23.2)	178 (31.6)	264 (34.5)	1088 (40.7)
Female	3025 (64)	557 (76.8)	384 (68.2)	502 (65.5)	1582 (59.1)
Other[e]	6 (0.1)	0 (0)	1 (0.2)	0 (0)	5 (0.2)
Education, n (%)					
Obligatory school or less	256 (5.4)	12 (1.7)	21 (3.7)	28 (3.7)	195 (7.3)
Apprenticeship	1379 (29.2)	171 (23.6)	131 (23.3)	209 (27.3)	868 (32.4)
High-school graduation	502 (10.6)	69 (9.5)	47 (8.3)	85 (11.1)	301 (11.3)
University or college	2526 (53.4)	466 (64.3)	358 (63.6)	434 (56.7)	1268 (47.4)
I do not know	66 (1.4)	7 (1)	6 (1.1)	10 (1.3)	42 (1.6)
Health literacy, n (%)					
Low health literacy	219 (4.6)	52 (7.2)	33 (5.9)	74 (9.7)	344 (12.9)
High health literacy	4510 (95.4)	673 (92.8)	529 (94.1)	692 (90.3)	2326 (87.1)
Employment[f], n (%)					
Full-time work	1707 (42.6)	—[g]	225 (40.0)	347 (45.3)	1135 (42.4)
Part-time work	674 (16.8)	—	102 (18.1)	136 (17.8)	436 (16.3)
Housewife and husband	125 (3.1)	—	18 (3.2)	24 (3.1)	83 (3.1)
Self-employed	303 (7.6)	—	37 (6.6)	46 (6.0)	220 (8.2)
Student	241 (6)	—	31 (5.5)	72 (9.4)	138 (5.2)
Employment status, n (%)					
Unemployed and currently looking for a job	133 (3.3)	—	23 (4.1)	19 (2.5)	91 (3.4)
Unemployed and not currently seeking employment	57 (1.4)	—	8 (1.4)	14 (1.8)	35 (1.3)
Incapacity	129 (3.2)	—	11 (2)	26 (3.4)	92 (3.4)
Retired	619 (15.5)	—	105 (18.7)	81 (10.6)	433 (16.2)
Unknown	15 (0.4)	—	3 (0.5)	1 (0.1)	11 (0.4)

[a]Survey 1: April 17 to May 14, 2020.

[b]Survey 2: May 15 to June 22, 2020.

[c]Survey 3: October 30 to December 1, 2020.

[d]Survey 4: June 18, 2021, to December 30, 2021.

[e]These were excluded from the regression analyses.

[f]Employment data were not collected during the first survey.

[g]Not available.

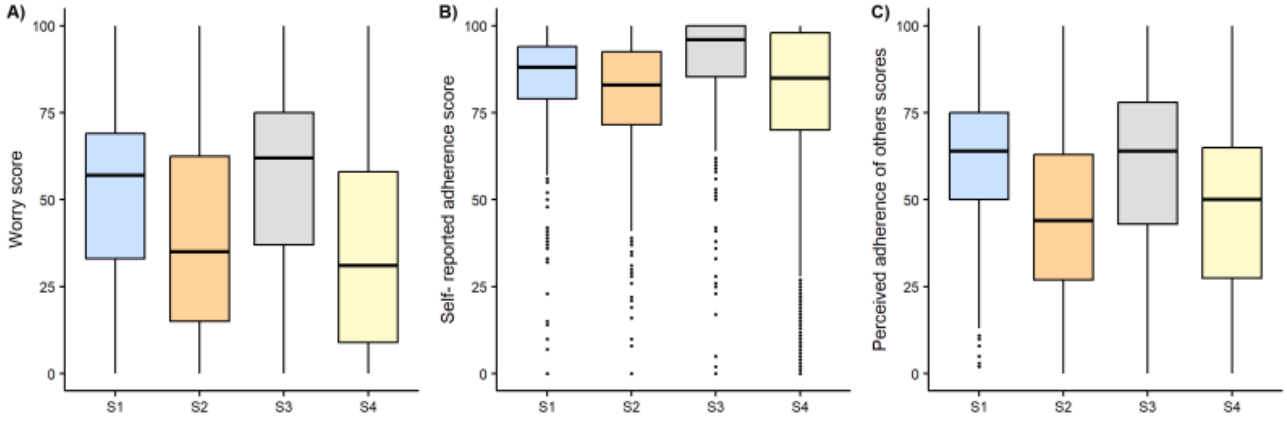
XSL•FO
RenderX

Main Results

Self-Reported Worry

Overall, the mean level of self-reported worry was 42.0% (SD 28.9). Upon univariate regression, self-reported worry levels differed significantly across surveys, with significantly higher levels in survey 1 (52%, 95% CI 50-54) and survey 3 (56%, 95% CI 54-58; Figure 2). Upon multivariable regression, the female gender was associated with a 4-point increase in level of worry (95% CI 2-5 points). Worry levels were 2 (95% CI 1-4) and 4 (95% CI 1-6) points lower among respondents aged 40 to 64 and over 64, respectively, when compared with the 18- to 39-year group. Higher health literacy was associated with a 3-point lower worry level (95% CI –6 to –1). Education was not associated with significant changes in self-reported level of worry (Table S5 in Multimedia Appendix 5).

Figure 2. Boxplots on a 0-100 scale of (A) worry, (B) self-reported adherence, and (C) perceived adherence of others to restrictions during four COVID-19 pandemic periods (April 17, 2020, to December 30, 2021), Switzerland. S: survey; survey 1: April 17 to May 14, 2020; survey 2: May 15 to June 22, 2020; survey 3: October 30 to December 1, 2020; survey 4: June 18, 2021, to December 30, 2021.



Self-Reported Adherence to Restrictions

Overall, respondents evaluated their own adherence to government restrictions at 81.4% (SD 21.1). Self-reported adherence was significantly (P<.001) higher in survey 1 (mean 84.8%, SD 14.2) and survey 3 (mean 89.6%, SD 15.5; Figure 2). A 2.9-point increase in the worry score was associated with a 10-point increase in self-reported adherence (95% CI 2.5-3.2; P<.001) after adjusting for the pandemic period, gender, age, education, and health literacy. These effects were more pronounced in women and older participants (Table S5 in Multimedia Appendix 5). Moreover, both older age categories (40-64 years and more than 64 years) were associated with a 7-point higher self-reported adherence than in the 18-39 years group (95% CI 6-8 for 40- to 64-year-olds and 6-9 for >64-year-olds). Higher health literacy was associated with a 4-point increase in self-reported adherence (95% CI 2-6) while the educational level was not (Table S5 in Multimedia Appendix 5).

Perceived Adherence of Others to Restrictions

Overall, respondents evaluated the adherence of others to government restrictions at 50.4% (SD 24.5%). Evaluated adherence did not differ between survey 1 and survey 3 nor between survey 2 and survey 4 but was significantly higher (P<.001) in survey 1 (mean 60.1%, SD 20.0%) and survey 3 (59.6%, SD 24.6%) than in survey 2 (45.1%, SD 22.5%) and survey 4 (46.2%, SD 24.5%; Figure 2). Adjusting for participant characteristics, age groups 40-64 years and >64 years were associated with, respectively, a 5- or 7-point higher perceived adherence of others (95% CI 3-6 and 5-9, respectively). A 10-point increase in worry level was associated with a 1-point decrease in perceived adherence of others to restrictions (95% CI –0.9 to –0.4). Higher education level was associated with a 1.5-point higher perceived adherence of others (95% CI, 0-3), whereas gender and health literacy were not (Table S5 in Multimedia Appendix 5).

Changes to the Daily Lives of Respondents

During the first period (survey 1), most respondents had experienced changes in their daily life (Table S6 in Multimedia Appendix 6). Later (surveys 2-4), 35% to 49% had either lost their job or had to close their business and 21% to 34% had lost part of their income. Respondents also reported feeling isolated, lonelier, and less productive during surveys 2 to 4. In decreasing order of importance, concerns during survey 3 were for “vulnerable people,” “living conditions,” the “economy,” “self and family,” “working conditions,” and the “possibility of another wave.” During survey 4, these concerns were similar but generally rated lower. Interestingly, “deterioration of working conditions” moved up from fifth to third rank, which had been “self and family” in survey 3 (Multimedia Appendix Table S7 in Multimedia Appendix 7 and Table S8 in Multimedia Appendix 8).

Discussion

Principal Findings

We conducted online surveys during different phases of the COVID-19 pandemic to describe the evolution of worry levels and to assess how these were associated with adherence to government restrictions. In Switzerland, the self-reported worry was highest during the first and second pandemic waves, corresponding to survey 1 and survey 3, at times of many COVID-19–related hospitalizations and deaths, and when a vaccine was not yet available. Women and younger people

XSL•FO
RenderX

reported higher levels of worry than men and older people. Education did not influence worry levels, while lower health literacy was associated with higher worry. Higher worry levels were associated with higher self-reported and perceived adherence of others to federal restrictions.

We found elevated worry levels during the more dramatic phases of the pandemic. In a systematic review covering 204 countries in 2020, higher anxiety was associated with higher COVID-19 incidence [23], a finding that was confirmed by Salanti et al [13]. In March 2020, Fitzpatrick et al [24] found that in a national sample in the United States, worry was higher in the regions with high COVID-19 incidence. In Ontario, Canada, COVID-19–related worry in young persons also increased during the early phases of the pandemic and then again in the autumn of 2020, when the incidence was higher [25]. So, anxiety and worry varied during the pandemic and increased repeatedly with the rising incidence of COVID-19. This is in line with findings on decreased mental health on a larger scale during the pandemic [11,26].

In periods of increased worry, we found higher self-reported adherence to government restrictions. Similarly, a study in Saudi Arabia described an association between higher anxiety levels and preventive practices among health care workers [27]. Another study identified fear as a predictor of behavioral change [28]. The association between worry and adherence in our study could indicate that worry was not overwhelming and that citizens felt in control of risks by respecting restrictions. We cannot exclude that this might have been different with higher anxiety levels.

Considering population subgroups, young adults were often the most anxious despite being less at risk of hospitalization or death [17,18,23,29]. Young people were worried about social isolation and develop depressive symptoms during school closings [30]. In Switzerland, students were concerned about whether they would be able to finish the 2020 university year [31], and lockdowns as well as their socioeconomic consequences were stressful for students [15]. Apart from concerns about the future, young people were not only worried about their own health but also about that of relatives. For example, in a study in Zurich, Switzerland, students were more concerned about the health of their parents and grandparents than their own [32]. Also, our finding of higher worry levels in women echoes several publications [17-19,33]. General factors potentially contributing to worry were the increasing risk of unemployment or loss of income, as well as loneliness and feeling less productive. One-third of respondents had lost part of their income by surveys 2 and 3 (31%, and 34%, respectively), with a slightly better situation in survey 4 (21%). The reported feelings of isolation, loneliness, and being less productive could further contribute to worry in some respondents. For example, studying for exams through online classes only, without any campus life, can be a source of worry compared with when stress from studying and exams is compensated by in-person interactions with teachers and colleagues. Four years after the pandemic, Sayed et al. [34] insist on the importance of addressing mental health of children and young adults during global crises and of recognizing long-term impacts. They further emphasize the need for research and public health prioritization of these important topics.

Overall, our findings are in line with publications that highlight the importance of addressing the many individual and collective aspects that influenced mental health during the COVID-19 pandemic, such as isolation, loneliness, and fear [35]. Many individuals demonstrated remarkable resilience, allowing society to avoid a general increase in loneliness [36]. However, population estimates may mask individual heterogeneity; loneliness is indeed a major public health concern and must be considered as a negative determinant of health [35]. Even though the pandemic is over, we must not forget its long-term effects on mental health and public health authorities should consider the differing impact of governmental decisions on the general population versus on individuals with pre-existing mental health conditions [37]. Our results of the worries of citizens and adherence to pandemic measures can be useful in preparing for future pandemics, for example, in considering criteria for and potential impact of restrictions on different subgroups of the population.

Strengths and Limitations

The strengths of our study were repeated surveys with similar questions and our relatively large sample size, allowing us to examine subgroups of the population.

Concerning limitations, data were collected through the online distribution of surveys in a simple, cost-effective, and feasible way during the rapidly evolving pandemic. Participation was more attractive to women and younger people with higher literacy and education. Also, participation was variable during data collection periods. The participant sample was more representative in the last and longest period (survey 4). This selection and variation need to be considered in the interpretation of our data which are prone to desirability, information, and selection bias. Different distribution channels and methods are probably needed for disadvantaged populations, as we showed in a recent study using our survey in a population of refugees and migrants [38]. Finally, another inherent limitation of our anonymous data collection is that we could not follow a cohort of persons throughout the pandemic.

Conclusion

Worry reached moderate levels and varied with COVID-19 incidence during the pandemic. Higher worry levels were associated with increased self-reported and perceived adherence of others to government restrictions. Younger people and women reported higher worry levels. Authorities should take population worry levels into account in planning and designing pandemic communication. Adapting communication to population subgroups should be considered for future health crises.

XSL•FO
RenderX

Acknowledgments

We would like to thank our community partners for distributing our electronic survey, specifically Yannis Papa Daniel from the Fédération Romande des Consommateurs and Dr Alix Miauton for the Coronacheck and Coronavax websites. We would also like to thank everyone who answered the surveys.

There was no specific financial support for the completion of this study.

Data Availability

The datasets generated during and/or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

CDL and VK wrote most of the manuscript. CM did most of the statistical analyses and wrote part of the article. SB contributed to analyzing and interpreting data and did part of the statistical analyses. M-AD contributed to designing the study and the questionnaires, collected data, and edited the manuscript. AG contributed to the logistics of the research team, interpreted the data, and edited the manuscript. KS contributed to designing the study and the questionnaires, collected data, and did statistical analyses. CVP was the instigator of the study and contributed to the study and questionnaire design, data collection, and manuscript writing. All authors revised the manuscript for intellectual content and approved it before submission. The authors indicate that they did not use generative artificial intelligence when writing their manuscript.

Conflicts of Interest

M-AD has contributed to the development of Option Grid patient decision aids and receives consulting income from EBSCO Health and royalties. KS receives salary support from the Leenaards Foundation. EBSCO Information Services sells subscription access to Option Grid patient decision aids.

Multimedia Appendix 1
Survey 1.
[PDF File (Adobe PDF File), 362 KB - ijmr_v14i1e55636_app1.pdf]

Multimedia Appendix 2
Survey 2.
[PDF File (Adobe PDF File), 343 KB - ijmr_v14i1e55636_app2.pdf]

Multimedia Appendix 3
Survey 3.
[PDF File (Adobe PDF File), 384 KB - ijmr_v14i1e55636_app3.pdf]

Multimedia Appendix 4
Survey 4.
[PDF File (Adobe PDF File), 370 KB - ijmr_v14i1e55636_app4.pdf]

Multimedia Appendix 5
Linear regression results for three models; all on visual analogue scales from 0 to 100 (0=not at all; 100=in all situations).
[PDF File (Adobe PDF File), 77 KB - ijmr_v14i1e55636_app5.pdf]

Multimedia Appendix 6
Perceived changes of daily life during the first pandemic wave.
[PDF File (Adobe PDF File), 45 KB - ijmr_v14i1e55636_app6.pdf]

Multimedia Appendix 7
Impact of restrictions on daily life, S2, S3, S4.
[PDF File (Adobe PDF File), 40 KB - ijmr_v14i1e55636_app7.pdf]

Multimedia Appendix 8
Concerns of respondents, S3, S4.
[PDF File (Adobe PDF File), 40 KB - ijmr_v14i1e55636_app8.pdf]

XSL•FO
RenderX

References

1. WHO. Listings of WHO's response to COVID-19. URL: https://www.who.int/news/item/29-06-2020-covidtimeline [accessed 2023-12-11]
2. Recommendations to Member States to improve hand hygiene practices to help prevent the transmission of the COVID-19 virus: interim guidance, 1 April 2020. Geneva, Switzerland: World Health Organization; 2020. URL: https://iris.who.int/handle/10665/331661
3. World Health Organization. Considerations for quarantine of contacts of COVID-19 cases: interim guidance, 25 June 2021. Geneva, Switzerland: World Health Organization; 2021. URL: https://iris.who.int/handle/10665/342004
4. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the Health Belief Model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
5. Zewdie A, Mose A, Sahle T, Bedewi J, Gashu M, Kebede N, et al. The health belief model's ability to predict COVID-19 preventive behavior: a systematic review. SAGE Open Med 2022;10::20503121221113668 [FREE Full text] [doi: 10.1177/20503121221113668] [Medline: 35898953]
6. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
7. Tannenbaum MB, Hepler J, Zimmerman RS, Saul L, Jacobs S, Wilson K, et al. Appealing to fear: a meta-analysis of fear appeal effectiveness and theories. Psychol Bull 2015;141(6):1178-1204 [FREE Full text] [doi: 10.1037/a0039729] [Medline: 26501228]
8. Bigsby E, Albarracín D. Self- and response efficacy information in fear appeals: a meta-analysis. J Commun 2022;72(2):241-263. [doi: 10.1093/joc/jqab048]
9. Lunn PD, Belton CA, Lavin C, McGowan FP, Timmons S, Robertson DA. Using behavioral science to help fight the coronavirus. JBPA 2020;3(1) [FREE Full text] [doi: 10.30636/jbpa.31.147]
10. Witte K, Allen M. A meta-analysis of fear appeals: implications for effective public health campaigns. Health Educ Behav 2000;27(5):591-615. [doi: 10.1177/109019810002700506] [Medline: 11009129]
11. Salari N, Hosseinian-Far A, Jalali R, Vaisi-Raygani A, Rasoulpoor S, Mohammadi M, et al. Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: a systematic review and meta-analysis. Global Health 2020;16(1):57 [FREE Full text] [doi: 10.1186/s12992-020-00589-w] [Medline: 32631403]
12. Santabárbara J, Lasheras I, Lipnicki DM, Bueno-Notivol J, Pérez-Moreno M, López-Antón R, et al. Prevalence of anxiety in the COVID-19 pandemic: an updated meta-analysis of community-based studies. Prog Neuropsychopharmacol Biol Psychiatry 2021;109:110207 [FREE Full text] [doi: 10.1016/j.pnpbp.2020.110207] [Medline: 33338558]
13. Salanti G, Peter N, Tonia T, Holloway A, White IR, Darwish L, MHCOVID Crowd Investigators, MHCOVID Crowd Investigators†. The impact of the COVID-19 pandemic and associated control measures on the mental health of the general population : a systematic review and dose-response meta-analysis. Ann Intern Med 2022;175(11):1560-1571 [FREE Full text] [doi: 10.7326/M22-1507] [Medline: 36252247]
14. Yong X, Zhang L. COVID-19 worry and mental health among the economically active population in Guangdong, China. Front Public Health 2022;10:882177 [FREE Full text] [doi: 10.3389/fpubh.2022.882177] [Medline: 35586016]
15. Shanahan L, Steinhoff A, Bechtiger L, Murray AL, Nivette A, Hepp U, et al. Emotional distress in young adults during the COVID-19 pandemic: evidence of risk and resilience from a longitudinal cohort study. Psychol Med 2022;52(5):824-833 [FREE Full text] [doi: 10.1017/S003329172000241X] [Medline: 32571438]
16. De Quervain DJF, Aerni A, Amini E, Bentz D, Coynel D, Gerhards C, et al. The Swiss Corona Stress Study Internet. Open Science Framework 2020 [FREE Full text] [doi: 10.31219/osf.io/jqw6a]
17. Moghanibashi-Mansourieh A. Assessing the anxiety level of Iranian general population during COVID-19 outbreak. Asian J Psychiatr 2020;51:102076 [FREE Full text] [doi: 10.1016/j.ajp.2020.102076] [Medline: 32334409]
18. Mazza C, Ricci E, Biondi S, Colasanti M, Ferracuti S, Napoli C, et al. A nationwide survey of psychological distress among Italian people during the COVID-19 pandemic: immediate psychological responses and associated factors. Int J Environ Res Public Health 2020;17(9):3165 [FREE Full text] [doi: 10.3390/ijerph17093165] [Medline: 32370116]
19. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
20. Selby K, Durand MA, Gouveia A, Bosisio F, Barazzetti G, Hostettler M, et al. Citizen responses to government restrictions in Switzerland during the COVID-19 pandemic: cross-sectional survey. JMIR Form Res 2020;4(12):e20871 [FREE Full text] [doi: 10.2196/20871] [Medline: 33156809]
21. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet E-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
22. Chew LD, Griffin JM, Partin MR, Noorbaloochi S, Grill JP, Snyder A, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med 2008;23(5):561-566 [FREE Full text] [doi: 10.1007/s11606-008-0520-5] [Medline: 18335281]

XSL•FO
RenderX

23. 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;398(10312):1700-1712 [FREE Full text] [doi: 10.1016/S0140-6736(21)02143-7] [Medline: 34634250]
24. Fitzpatrick KM, Drawve G, Harris C. Facing new fears during the COVID-19 pandemic: the state of America's mental health. J Anxiety Disord 2020;75:102291 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102291] [Medline: 32827869]
25. Hawke LD, Szatmari P, Cleverley K, Courtney D, Cheung A, Voineskos AN, et al. Youth in a pandemic: a longitudinal examination of youth mental health and substance use concerns during COVID-19. BMJ Open 2021;11(10):e049209 [FREE Full text] [doi: 10.1136/bmjopen-2021-049209] [Medline: 34716160]
26. Mental Health and COVID-19: early evidence of the pandemic's impact: Scientific brief, 2 March 2022. Geneva, Switzerland: World Health Organization; 2022. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1 [accessed 2024-12-27]
27. Alshagrawi S, Alhowti S. Knowledge, anxiety, and practice during the COVID-19 crisis among HCWs in Saudi Arabia. J Educ Health Promot 2022;11:384 [FREE Full text] [doi: 10.4103/jehp.jehp_308_22] [Medline: 36618461]
28. Harper CA, Satchell LP, Fido D, Latzman RD. Functional fear predicts public health compliance in the COVID-19 pandemic. Int J Ment Health Addict 2021;19(5):1875-1888 [FREE Full text] [doi: 10.1007/s11469-020-00281-5] [Medline: 32346359]
29. Hyland P, Shevlin M, McBride O, Murphy J, Karatzias T, Bentall RP, et al. Anxiety and depression in the republic of Ireland during the COVID-19 pandemic. Acta Psychiatr Scand 2020;142(3):249-256. [doi: 10.1111/acps.13219] [Medline: 32716520]
30. Hawes MT, Szenczy AK, Klein DN, Hajcak G, Nelson BD. Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic. Psychol Med 2022;52(14):3222-3230 [FREE Full text] [doi: 10.1017/S0033291720005358] [Medline: 33436120]
31. Amendola S, von Wyl A, Volken T, Zysset A, Huber M, Dratva J. A longitudinal study on generalized anxiety among university students during the first wave of the COVID-19 pandemic in Switzerland. Front Psychol 2021;12:643171 [FREE Full text] [doi: 10.3389/fpsyg.2021.643171] [Medline: 33776867]
32. Dratva J, Zysset A, Schlatter N, von Wyl A, Huber M, Volken T. Swiss university students' risk perception and general anxiety during the COVID-19 pandemic. Int J Environ Res Public Health 2020;17(20):7433 [FREE Full text] [doi: 10.3390/ijerph17207433] [Medline: 33066008]
33. Robinson E, Sutin AR, Daly M, Jones A. A systematic review and meta-analysis of longitudinal cohort studies comparing mental health before versus during the COVID-19 pandemic in 2020. J Affect Disord 2022;296:567-576 [FREE Full text] [doi: 10.1016/j.jad.2021.09.098] [Medline: 34600966]
34. Sayed AA, El-Gendy AA, Aljohani AK, Haddad RA, Taher OH, Senan AM. The effects of COVID-19 on the mental health of children and adolescents: a review. Cureus. 2024. URL: https://www.cureus.com/articles/207363-the-effects-of-covid-19-on-the-mental-health-of-children-and-adolescents-a-review [accessed 2024-06-13]
35. Lederman Z. Loneliness at the age of COVID-19. J Med Ethics 2023;49(9):649-654. [doi: 10.1136/jme-2022-108540] [Medline: 36600628]
36. Luchetti M, Lee JH, Aschwanden D, Sesker A, Strickhouser JE, Terracciano A, et al. The trajectory of loneliness in response to COVID-19. Am Psychol 2020;75(7):897-908 [FREE Full text] [doi: 10.1037/amp0000690] [Medline: 32567879]
37. Maher A. Tracing the bounds of distress: mental health and the Lancet commission on lessons for the future from the COVID-19 pandemic. J Glob Health 2024;14:03022 [FREE Full text] [doi: 10.7189/jogh.14.03022] [Medline: 38751302]
38. Morisod K, Durand MA, Selby K, Le Pogam MA, Grazioli VS, Sanchis Zozaya J, et al. Asylum seekers' responses to government COVID-19 recommendations: a cross-sectional survey in a Swiss Canton. J Immigr Minor Health 2023;25(3):570-579 [FREE Full text] [doi: 10.1007/s10903-022-01436-3] [Medline: 36508030]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet e-Surveys
REDCap: Research Electronic Data Capture

Edited by T de Azevedo Cardoso; submitted 22.12.23; peer-reviewed by S Haile, JV Sánchez-Ortí; comments to author 23.05.24; revised version received 13.06.24; accepted 25.09.24; published 07.01.25.

Please cite as:
Kraege V, Dumans-Louis C, Maglieri C, Bochatay S, Durand MA, Garnier A, Selby K, von Plessen C
Citizen Worry and Adherence in Response to Government Restrictions in Switzerland During the COVID-19 Pandemic: Repeated Cross-Sectional Online Surveys
Interact J Med Res 2025;14:e55636
URL: https://www.i-jmr.org/2025/1/e55636
doi:10.2196/55636
PMID:

XSL•FO
RenderX

©Vanessa Kraege, Céline Dumans-Louis, Céline Maglieri, Séverine Bochatay, Marie-Anne Durand, Antoine Garnier, Kevin Selby, Christian von Plessen. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 07.01.2025. 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

French Versions of 4 English Questionnaires on Problematic Smartphone Use: Cross-Cultural Linguistic Translation and Adaptation Study

Islam El Boudi[1], MEng; Mathilde Riant[1], PhD; Alexandre Bellier[1,2], MD, PhD; Nicolas Vuillerme[1,3], PhD

[1]Faculty of Medicine, AGEIS, Grenoble Alpes University, Grenoble, France

[2]Inserm CIC 1406, Grenoble Alpes University Hospital (CHUGA), Grenoble Alpes University, 38000 Grenoble, France

[3]Institut Universitaire de France, 75005 Paris, France

Corresponding Author:
Islam El Boudi, MEng
Faculty of Medicine
AGEIS, Grenoble Alpes University
Jean Roget Bldg, 3rd Fl.
Grenoble, 38000
France
Phone: 33 695685616
Email: islam.el-boudi@univ-grenoble-alpes.fr

Abstract

Background: Excessive use of smartphones is recognized as a major problem in our modern society and can have dramatic consequences on the health of adolescents and young adults. Measuring problematic smartphone use in research and clinical practice is generally operationalized with self-reported questionnaires. In order to comprehensively assess the issue of problematic smartphone usage within the French population, it is imperative to employ validated French-language questionnaires. However, at this point, existing questionnaires are primarily available in English. Furthermore, to the best of our knowledge, these English questionnaires have yet to undergo validation processes for French-speaking cohorts.

Objective: The aim of this study was to perform a cross-cultural translation of the Smartphone Addiction Scale, Nomophobia Questionnaire, Problematic Use of Mobile Phones scale, and Smartphone Addiction Proneness Scale to French.

Methods: The translation process was performed using the forward/backward method. The first translation phase involved asking 4 independent French translators to translate the original English version of the questionnaires into French. In the second phase, the French version was backtranslated to English by a native English speaker. In the third phase, 2 concept experts were asked to comment and suggest modifications to the statements if necessary. Finally, the last version of the translated questionnaires was presented to 18 participants to assess the clarity, intelligibility, and acceptability of the translations.

Results: During the forward translation step, the translation differences were minor. During the backward translation, the English native speaker correctly backtranslated 18 of the 33 items of the Smartphone Addiction Scale, 17 of the 20 items of the Problematic Use of Mobile Phones scale, and 13 of the 15 items of the Smartphone Addiction Proneness Scale. Backtranslation for the Nomophobia Questionnaire was less satisfactory, with only 10 out of 20 items that were correctly backtranslated. The linguistic verification step revealed a minimal modification for the 4 questionnaires. The participants also suggested few improvements that we have considered for the final version. We produced the final version directly after this step.

Conclusions: We successfully adapted and effectively translated 4 questionnaires that assess problematic smartphone use to French. This step is a prerequisite for the validation of the French questionnaires. These adapted measures can serve as valuable research instruments for investigating and addressing issues related to problematic smartphone use in French-speaking countries and for making international comparisons.

(Interact J Med Res 2025;14:e53958) doi:10.2196/53958

KEYWORDS

problematic use; smartphone; French questionnaire; linguistic translation; forward/backward process; mobile phone

XSL•FO
RenderX

Introduction

Today, the excessive use of smartphones and screens, in general, is recognized as a major problem in our society. Recent world events that have led to individuals being confined to their homes for several months have intensified the amount of scientific and political concerns surrounding this issue. Excessive smartphone use can have dramatic consequences for the health and cognitive development of adolescents and young adults [1]. Indeed, difficulty in regulating smartphone use can lead to serious pathological disorders such as addiction to video games [2].

Although health care professionals, researchers, and politicians recognize that excessive smartphone use is problematic, its legitimacy as a pathological disorder is not yet recognized by the American Psychiatric Association [2]. For this reason, the notion of problematic smartphone use is favored by a large number of researchers. Problematic smartphone use does, however, refer to symptoms of dependence such as loss of control, overuse, increased tolerance (ie, spending more time on the smartphone to be satiated), or withdrawal symptoms once individuals are no longer in possession of their smartphones [3].

In order to identify individuals at risk of problematic use, the measurement of smartphone use practices is generally operationalized in research by using self-reported questionnaires. Three very recently published systematic reviews [3-5] identified 4 measurement scales most commonly used to measure problematic smartphone use, 3 of which are used with young adult students, that is, the Smartphone Addiction Scale [6], the Nomophobia Questionnaire [7], and the Problematic Use of Mobile Phones scale [8], and 1 for adolescents, that is, the Smartphone Addiction Proneness Scale [9]. What these 4 scales have in common is that they assess problematic smartphone use across several dimensions, often inspired by the fourth and fifth versions of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV and DSM-V) [2,10].

The Smartphone Addiction Scale was designed by Kwon et al [6] based on the diagnostic criteria for internet gaming addiction in DSM-IV [10]. It measures problematic smartphone use along 6 dimensions: disruption of daily life due to smartphone use, positive anticipation of use, withdrawal symptoms, overuse, tolerance, and importance of the virtual world for social relationships. This scale has the advantage of being validated in several languages, including English [6], Arabic [11], Brazilian [12], Malaysian [13], and Persian [14]. It also has good cross-cultural consistency, with Cronbach α ranging from 0.93 to 0.97. Although the Smartphone Addiction Scale has been translated and validated in French in a shorter 10-statement version [15], the original 33-items scale has not been translated and validated in French.

The Nomophobia Questionnaire was designed by Yildirim and Correia [7] and measures individuals' fear of being separated from their smartphones. It measures 4 dimensions of smartphone separation fear: not being able to communicate, losing connection with the outside world, not being able to access information, and giving up comfort. According to the research work of Jahrami et al [16], this questionnaire has been translated into 7 languages (eg, Chinese [17], Italian [18]) and consistently has very good internal consistency ($0.88 \leq \alpha \leq 0.96$).

The Problematic Use of Mobile Phones scale, designed by Merlo et al [8], draws heavily on the criteria for online gaming addiction in DSM versions IV [10] and V [2]. This scale considers 10 dimensions of problematic smartphone use in young adults: tolerance, withdrawal symptoms, smartphone use longer than expected, time spent on the smartphone, irrepressible desire to use the smartphone, abandonment or reduction of other activities, smartphone use despite physical and psychological consequences, smartphone use in risky situations, smartphone use despite social consequences, and inability to fulfill obligations due to use. Although being less validated, this scale has been translated into English [8], German [19], and Arabic [20], and its consistency appears very satisfactory ($0.90 \leq \alpha \leq 0.94$).

To date, 2 other scales measuring problematic smartphone use by young adults have been validated in French. The first scale, that is, the Implicit Association Test (smartphone) [21] is strongly inspired by the internet addiction test developed by Young [22] and not according to a set of diagnostic criteria for problematic smartphone use. However, the factor structures and internal consistency of the scale ($\alpha=0.93$) were good. The other scale, that is, the Problematic Mobile Phone Use Questionnaire by Lopez-Fernandez et al [23] is divided into 3 categories and has good internal consistency: dangerous use ($\alpha=0.81$), prohibited use ($\alpha=0.74$), and dependent use ($\alpha=0.90$). The strength of this scale lies in its cross-cultural validity (eg, French, German, Hungarian, English, Finnish, Italian, Polish, Spanish), but surprisingly, it is one of the few scales that is not based on any version of DSM. However, recent research has shown that taking inspiration from gaming disorder criteria to measure problematic smartphone use is a way to better understand this phenomenon [24].

The Smartphone Addiction Proneness Scale by Kim et al [9] is inspired by the assessment criteria of Young's [22] Internet Addiction Test scale. To our knowledge, it is one of the few scales designed to measure problematic smartphone use in adolescents. The Smartphone Addiction Proneness Scale assesses 4 main dimensions: craving symptoms, tolerance, predominance of virtual life, and disruption of coping in daily life. To date, it has been translated into Korean (ie, translated from English) [9], Malaysian [25], German [26], and Chinese [27]. The interrater internal consistency of this Korean scale is consistently acceptable ($\alpha=0.88$).

Given the importance of completing proposals for measuring problematic smartphone use based primarily on the findings of DSM and to better understand this societal issue, the aim of this study is to linguistically translate the Smartphone Addiction Scale, the Nomophobia Questionnaire, the Problematic Use of Mobile Phones scale, and the Smartphone Addiction Proneness Scale to French so that they are conceptually equivalent to their original versions. The cross-cultural adaptation process will be performed using the forward/backward method, a commonly used method [28-32] and recommended by the World Health Organization [33]. The translation will be considered valid when the tests performed by the participants are considered conclusive.

XSL•FO
RenderX

This step is a prerequisite for the quantitative validation of the questionnaires.

Methods

Background

To assess the conceptual equivalence of the questionnaires, that is, whether the items in the original language have a similar meaning to the French version, we opted for the forward/backward translation method. This is the most commonly used technique for cross-cultural research, and we followed the recommendations of Epstein et al [34]. We have decided to validate these instruments for use exclusively in France.

Ethics Approval

This study was approved by the Grenoble Alpes University Hospital Center and has received ethics approval from the South-East I Ethics Committee for the Protection of Individuals (approval 2022-A01943-40).

Questionnaires

The 4 questionnaires targeted for the cross-cultural procedure were as follows. The Smartphone Addiction Scale [6] contains 33 items, and the response scale ranges from 1 (strongly disagree) to 6 (strongly agree). Each participant's total score can range from 33 to 188 points, and a higher score indicates more problematic smartphone use. The Nomophobia Questionnaire [7] contains 20 items, and the response scale ranges from 1 (strongly disagree) to 7 (strongly agree); each participant's total score can therefore vary from 20 to 140 points, and a higher score indicates more problematic smartphone use. The Problematic Use of Mobile Phones scale contains 20 items [8]. The response scale ranges from 1 (not at all in agreement) to 5 (totally in agreement), and each participant's total score can vary from 20 to 100 points, and a higher score indicates more problematic smartphone use. Finally, the Smartphone Addiction Proneness Scale by Kim et al [9] contains 15 items. The response scale ranges from 1 (not at all in agreement) to 4 (totally in agreement), and each participant's total score can vary from 15 to 60 points, and a higher score indicates more problematic smartphone use.

Forward Translation

The first translation phase involved recruiting independent translators who were both 2 native French speakers bilingual in English and 2 native English speakers bilingual in French and asking them to translate the original English version into French. It was recommended that at least one of the translators knows the concept of the questionnaires to be measured and that at least one of them does not know the objective of the questionnaire. In this study, 4 qualified translators were involved: 2 translators were familiar with the concept of problematic smartphone use, while the other 2 were not. Two of the experts had backgrounds in health and psychology research: one was a novice and the other had prior experience in translation. At the end of the 4 translations, a single version was obtained after a reconciliation meeting of the 4 translators. Each questionnaire was checked for errors in spelling, grammar, punctuation, and translation of terminology and style against the original English version.

Backward Translation

In the second phase, the French versions were retranslated into English by a native English speaker who had no clinical or medical expertise. The instructions that were given to the native English speaker were to translate each questionnaire literally. By comparing the translated English versions with the original questionnaires, the French translations were modified to be consistent with the originals. This step ensured the accuracy of the translations and highlighted any phrasing that could cause confusion.

Linguistic Verification

In the third phase, 2 professionals familiar with the concept of problematic smartphone use were asked to comment and suggest modifications to the statements if necessary. We followed a universalist approach for equivalence, assessing conceptual, item, semantic, operational, measurement, and functional equivalence [35].

Final Verification

The final verification phase involved performing a preliminary pilot test with participants involved in the problem being measured to assess the clarity, intelligibility, and acceptability of the translations. Between December 2022 and June 2023, 18 French native speaker participants—10 students older than 18 years (ie, 6 girls and 4 boys) and 8 adolescents aged 12-17 years (ie, 6 girls and 2 boys)—completed the questionnaires for the final version during face-to-face interviews. Each interview lasted a maximum of 30 minutes. For minor participants, consent to take part in the study was obtained from the parents, while adult participants gave their own consent. Participants could be in good health or experiencing any medical condition and they could come from rural or urban areas in France. During the interviews, a researcher asked the participants questions such as "Did you understand the instructions, the items, and the response scales?"; "How did you interpret this item?"; or "What would you suggest as a reformulation?" The principal investigator checked the proofs of the final version and corrected any errors.

Data Analysis

To analyze the qualitative data from the translation process, we employed a thematic approach focused on assessing the conceptual equivalence of the translated items. Responses from the participants during the individual interviews were transcribed and systematically analyzed to identify elements that could pose comprehension issues. Particular attention was given to semantic discrepancies between the original and translated versions as well as suggestions for reformulation provided by the participants. The data were then coded according to the types of mistranslations identified (eg, item comprehension, appropriateness of terms used) and grouped into recurring themes. This approach enabled us to evaluate the clarity and acceptability of the translations as well as to adjust the problematic items.

XSL•FO
RenderX

Results

Characteristics of the Translators and the Participants

The age, gender, and specific cultural characteristics of the translators and the student and adolescent participants are detailed in Table 1.

Table 1. Age, gender, and cultural characteristics of the translators and the participants.

Translation type	Translators/participants (n)	Mean age (years)	Gender	Cultural characteristics
Forward translation	4 translators	25.5 (range 20-31)	2 females, 2 males	• 2 native English, 2 native French • 2 were familiar with the concept of problematic smartphone use (eHealth and psychology) • 2 were not familiar with the concept (economics and education) • 2 were from the south of France and 2 were from the north
Backward translation	1 translator	28	1 female	• Native English • No clinical or medical expertise
Linguistic verification	2 experts	27.5 (range 24-31)	1 female, 1 male	• 2 native French
Final verification				
	8 adolescents	14.5 (range 12-17)	6 females, 2 males	• Good health or experiencing any medical condition • 5 from rural areas and 3 from urban areas in France
	10 students	19.8 (range 18-24)	6 females, 4 males	• Good health or experiencing any medical condition • 6 students in health sciences, 3 students in psychology, 1 student in philosophy

Forward Translation

The 4 translators translated the titles, dimensions, instructions, response modalities, and statements of each questionnaire. These 4 translations were synthesized into a consensual version A. Concerning questionnaire dimensions, a few terms were more difficult to translate from English to French. For example, the dimension “withdrawal” of the Smartphone Addiction Scale, the Problematic Use of Mobile Phones scale, and the Smartphone Addiction Proneness Scale was translated as “sevrage.” For the Nomophobia Questionnaire, 1 dimension was more challenging to translate into French: “giving up convenience.” The term “convenience” can have several meanings in French, but its translation as “confort” (comfort) appears to be the most appropriate, especially when defining this dimension as the loss of the comfort provided by smartphone use. The concept of “craving” for the Nomophobia Questionnaire is particularly complex to define in French as it does not have a direct equivalent. In clinical terms, craving corresponds to a strong impulsive desire to use one’s object of addiction, which in our context is the respondent’s smartphone use. Therefore, in French, we have defined this dimension as “désir” (desire). For the other dimensions, the terms used in French were mostly similar.

For the items in the Smartphone Addiction Scale, rather than using the infinitive form of verbs as proposed in the original version, we translated them by systematically beginning the verb with the first-person singular conjugation. For example, for the original item “missing planned work due to smartphone use,” we translated it as “J’oublie du travail planifié à cause de l’utilisation du smartphone.” This will help participants better understand the statements. Three statements were more difficult to translate into French: “feeling pleasant or excited while using a smartphone” and “feeling impatient and fretful when I am not holding my smartphone.” The term “pleasant” was suggested in French as “calme” (calm), and since the term “impatient” does not have a direct equivalent in French, we propose the homograph “impatient.” The verb “to annoy” present in 2 items of the Nomophobia Questionnaire required more debate regarding its translation. This term can imply constraint, boredom, or irritation depending on the context; therefore, we chose to translate it as “contrariété” (contrariety) to preserve the statement and its dimension without distorting it. The statement “using a smartphone is more enjoyable than spending time with family or friends” from the Smartphone Addiction Proneness Scale also sparked discussions among translators because not everyone agreed on how to translate “enjoyable.” This term can signify a pleasant or amusing moment depending on the context. Here, we opted for the most common translation, “agréable” (pleasant). For the other items, the differences in translation were minor and did not alter the literal meaning of the statements.

Backward Translation

Following this first translation, a native English translator backtranslated the questionnaires from the consensus version A into an English version B. This version B was compared to the original questionnaire in English. For the Smartphone

XSL•FO
RenderX

Addiction Scale, 18 items were perfectly backtranslated, while 14 were slightly modified, but this did not alter their meaning and understanding. One item was retranslated, changing its meaning: “missing planned work due to smartphone use” was backtranslated as “I use my smartphone longer than planned.” We reworked this item in French to maintain its original meaning as “Je manque du travail planifié à cause de l’utilisation du smartphone.” For the Problematic Use of Mobile Phones scale, 17 statements were perfectly backtranslated, and the others underwent slight modifications without impacting their meaning and understanding. Backtranslation was even better for the Smartphone Addiction Proneness Scale, with 13 out of 15 items perfectly backtranslated and 2 having slight modifications without consequence. The backtranslation for the Nomophobia Questionnaire was satisfactory, with 10 out of 20 items varying slightly in wordings but without altering their meaning. The others were perfectly backtranslated. Thus, we obtained a translated version C for the 4 questionnaires.

Linguistic Verification

The 2 researchers responsible for linguistic verification submitted 4 suggestions for conceptual modifications for the Smartphone Addiction Scale, 3 for the Nomophobia Questionnaire, and 1 for the Problematic Use of Mobile Phones scale, some of which were incorporated into the final C version. No additional modifications were suggested for the Smartphone Addiction Proneness Scale.

Final Verification

In this final step, feedback from adolescents and students was minimal; therefore, we produced a final version D based on their input. Overall, they understood all the instructions, items, and response modalities well. Participants encountered no comprehension problems except for 3 items. For the Nomophobia Questionnaire, 2 students reported a comprehension difficulty with item 18. In English, it refers to checking “online connections and networks,” and we translated it as “online notifications and networks.” However, “online networks” made less sense than “social networks” for these students. We therefore modified the item. In addition, item 6 of the Problematic Use of Mobile Phones scale was difficult for one student to understand, as the translators had translated it by mixing the present and past tenses. We therefore modified the item from “J’ai pensé par le passé qu’il n’était pas normal de passer autant de temps sur le smartphone comme je le fais” to “Il m’est arrivé de penser qu’il n’était pas normal de passer autant de temps que moi à utiliser un smartphone.” In this way, students gained a better understanding of the item. Finally, item 22 of the Smartphone Addiction Scale had to be clarified, as 2 participants did not understand what “losing a friend” meant. In fact, the meaning of loss is important in French, as it could mean literally losing a friend through death or signify a friendship breakup. We have therefore clarified at the end of the item that this refers to friendship breakup in this context. The final translated versions are listed in Tables 2-5.

XSL•FO
RenderX

Table 2. Problematic Use of Mobile Phones scale translation.

Item number	Item description	Response methods				
		1[a]	2	3	4	5[b]
1	ET[c]: When I decrease the amount of time spent using my cell phone I feel less satisfied. FT[d]: Quand je réduis le temps passé sur mon smartphone, je me sens moins satisfait·e.					
2	ET: I need more time using my cell phone to feel satisfied than I used to need. FT: J'ai besoin de plus de temps d'utilisation de mon smartphone afin de me sentir aussi satisfait·e qu'auparavant.					
3	ET: When I stop using my cell phone, I get moody and irritable. FT: Quand j'arrête d'utiliser mon smartphone, je deviens de mauvaise humeur et irritable.					
4	ET: It would be very difficult emotionally to give up my cell phone. FT: Ce serait difficile émotionnellement de renoncer à mon smartphone.					
5	ET: The amount of time I spend using my cell phone keeps me from doing other important work. FT: Le temps passé sur mon smartphone m'empêche de faire d'autres tâches importantes.					
6	ET: I have thought in the past that it is not normal to spend as much time using a cell phone as I do. FT: Il m'est arrivé·e de penser qu'il n'était pas normal de passer autant de temps que moi à utiliser un smartphone.					
7	ET: I think I might be spending too much time using my cell phone. FT: Je pense que je passe trop de temps sur mon smartphone.					
8	ET: People tell me I spend too much time using my cell phone. FT: Les gens me disent que je passe trop de temps sur mon smartphone.					
9	ET: When I am not using my cell phone, I am thinking about using it or planning the next time I can use it. FT: Quand je n'utilise pas mon smartphone, je pense à l'utiliser ou à prévoir ma prochaine utilisation.					
10	ET: I feel anxious if I have not received a call or message in some time. FT: Je me sens anxieux·se si je n'ai pas reçu d'appel ou de message depuis un certain moment.					
11	ET: I have ignored the people I'm with in order to use my cell phone. FT: J'ai ignoré les personnes avec qui j'étais pour utiliser mon smartphone.					
12	ET: I have used my cell phone when I knew I should be doing work/schoolwork. FT: J'ai utilisé mon téléphone alors que je savais que je devais travailler/faire des devoirs.					
13	ET: I have used my cell phone when I knew I should be sleeping. FT: J'ai déjà utilisé mon smartphone alors que je savais que je devais dormir.					
14	ET: When I stop using my cell phone because it is interfering with my life, I usually return to it. FT: Quand j'arrête d'utiliser mon smartphone parce qu'il interfère avec ma vie, je finis généralement par l'utiliser à nouveau.					
15	ET: I have gotten into trouble at work or school because of my cell phone use. FT: J'ai eu des problèmes au travail ou à l'école à cause de l'utilisation de mon smartphone.					
16	ET: At times, I find myself using my cell phone instead of spending time with people who are important to me and want to spend time with me. FT: Parfois, je me retrouve à utiliser mon smartphone au lieu de passer du temps avec des personnes qui sont importantes pour moi et qui souhaitent passer du temps avec moi.					
17	ET: I have used my cell phone when I knew it was dangerous to do so. FT: J'ai utilisé mon téléphone alors que je savais qu'il était dangereux de le faire.					
18	ET: I have almost caused an accident because of my cell phone use. FT: J'ai presque causé un accident à cause de l'utilisation de mon smartphone.					
19	ET: My cell phone use has caused me problems in a relationship. FT: L'utilisation de mon smartphone a causé des problèmes dans une relation.					
20	ET: I have continued to use my cell phone even when someone asked me to stop. FT: J'ai continué d'utiliser mon smartphone, même quand quelqu'un m'a demandé d'arrêter.					

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: strongly agree; French translation: fortement en accord.

XSL•FO
RenderX

[c]ET: English translation.

[d]FT: French translation.

XSL•FO
RenderX

Table 3. Smartphone Addiction Scale translation.

Item number	Item description	Response methods					
		1[a]	2[b]	3[c]	4[d]	5[e]	6[f]
1	ET[g]: Missing planned work due to smartphone use. FT[h]: Je manque du travail planifié à cause de l'utilisation du smartphone.						
2	ET: Having a hard time concentrating in class while doing assignments or while working due to smartphone use. FT: J'ai des difficultés à me concentrer en classe, pendant mes devoirs ou pendant le travail à cause de l'utilisation du smartphone.						
3	ET: Experiencing lightheadedness or blurred vision due to excessive smartphone use. FT: Je ressens des vertiges ou une vision floue à cause de l'utilisation excessive du smartphone.						
4	ET: Feeling pain in the wrist or at the back of the neck while using a smartphone. FT: Je ressens de la douleur dans les poignets ou derrière le cou pendant que j'utilise un smartphone.						
5	ET: Feeling tired and lacking adequate sleep due to excessive smartphone use. FT: Je me sens fatigué et en manque de sommeil suffisant à cause de l'utilisation excessive du smartphone.						
6	ET: Feeling calm or cozy while using a smartphone. FT: Je me sens calme et réconforté·e quand j'utilise un smartphone.						
7	ET: Feeling pleasant or excited while using a smartphone. FT: Je me sens calme ou excité·e en utilisant le smartphone.						
8	ET: Feeling confident while using a smartphone. FT: Je me sens confiant·e en utilisant le smartphone.						
9	ET: Being able to get rid of stress with a smartphone. FT: Je suis capable de me débarrasser du stress avec un smartphone.						
10	ET: There is nothing more fun to do than using my smartphone. FT: Il n'y a rien de plus amusant que d'utiliser mon smartphone.						
11	ET: My life would be empty without my smartphone. FT: Ma vie serait vide sans mon smartphone.						
12	ET: Feeling most liberal while using a smartphone. FT: Je me sens libre quand j'utilise un smartphone.						
13	ET: Using a smartphone is the most fun thing to do. FT: Utiliser un smartphone est la chose la plus amusante à faire.						
14	ET: Won't be able to stand not having a smartphone. FT: Je ne supporterais pas de ne pas avoir de smartphone.						
15	ET: Feeling impatient and fretful when I am not holding my smartphone. FT: Je me sens impatient·e et irrité·e quand je ne tiens pas mon smartphone.						
16	ET: Having my smartphone in my mind even when I am not using it. FT: Je pense à mon smartphone même lorsque je ne l'utilise pas.						
17	ET: I will never give up using my smartphone even when my daily life is already greatly affected by it. FT: Je ne renoncerais jamais à utiliser mon smartphone, même si ma vie quotidienne en est déjà fortement affectée.						
18	ET: Getting irritated when bothered while using my smartphone. FT: Je me sens irrité·e quand on me dérange pendant que j'utilise mon smartphone.						
19	ET: Bringing my smartphone to the toilet even when I am in a hurry to get there. FT: J'apporte mon smartphone aux toilettes même lorsque je suis pressé·e d'y aller.						
20	ET: Feeling great meeting more people via smartphone use. FT: Je me sens bien en rencontrant plus de personnes par le biais de l'utilisation de mon smartphone.						

XSL•FO
RenderX

Item number	Item description	Response methods					
		1[a]	2[b]	3[c]	4[d]	5[e]	6[f]
21	ET: Feeling that my relationships with my smartphone buddies are more intimate than my relationships with my real-life friends. FT: J'ai l'impression que mes relations virtuelles sont plus intimes que mes relations avec mes amis de la vie réelle.						
22	ET: Not being able to use my smartphone would be as painful as losing a friend. FT: Ne pas pouvoir utiliser mon smartphone serait aussi douloureux que de perdre un ami (rupture amicale).						
23	ET: Feeling that my smartphone buddies understand me better than my real-life friends. FT: J'ai l'impression que mes amis virtuels me comprennent mieux que mes amis de la vie réelle.						
24	ET: Constantly checking my smartphone so as not to miss conversations between other people on Twitter or Facebook. FT: Je vérifie constamment mon smartphone afin de ne manquer aucune conversation entre d'autres personnes sur Twitter ou Facebook.						
25	ET: Checking social networking service sites like Twitter or Facebook right after waking up. FT: Je vérifie les réseaux sociaux comme Twitter ou Facebook juste après le réveil.						
26	ET: Preferring talking with my smartphone buddies to hanging out with my real-life friends or with the other members of my family. FT: Je préfère discuter avec mes amis virtuels plutôt que passer du temps avec mes amis de la vie réelle ou avec les membres de ma famille.						
27	ET: Preferring searching from my smartphone to asking other people. FT: Je préfère faire mes recherches à partir de mon smartphone plutôt que de demander à d'autres personnes.						
28	ET: My fully charged battery does not last for one whole day. FT: Ma batterie pleinement chargée ne dure pas une journée entière.						
29	ET: Using my smartphone longer than I had intended. FT: J'utilise mon smartphone plus longtemps que prévu.						
30	ET: Feeling the urge to use my smartphone again right after I stopped using it. FT: Je ressens le besoin d'utiliser à nouveau mon smartphone juste après avoir arrêté de l'utiliser.						
31	ET: Having tried time and again to shorten my smartphone use time but failing all the time. FT: J'ai essayé à plusieurs reprises de réduire le temps d'utilisation de mon smartphone mais j'échoue à chaque fois.						
32	ET: Always thinking that I should shorten my smartphone use time. FT: Je pense toujours que je devrais réduire mon utilisation du smartphone.						
33	ET: The people around me tell me that I use my smartphone too much. FT: Les gens autour de moi me disent que j'utilise trop mon smartphone.						

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: disagree; French translation: en désaccord.

[c]English translation: weakly disagree; French translation: un peu en désaccord.

[d]English translation: weakly agree; French translation: un peu en accord.

[e]English translation: agree; French translation: en accord.

[f]English translation: strongly agree; French translation: fortement en accord.

[g]ET: English translation.

[h]FT: French translation.

XSL•FO
RenderX

Table 4. Nomophobia Questionnaire translation.

Item number	Item description	Response methods						
		1[a]	2	3	4	5	6	7[b]
1	ET[c]: I would feel uncomfortable without constant access to information through my smartphone. FT[d]: Je me sentirais mal à l'aise sans l'accès constant à l'information au travers de mon smartphone.							
2	ET: I would be annoyed if I could not look information up on my smartphone when I wanted to do so. FT: Je me sentirais contrarié·e si je ne pouvais pas chercher de l'information sur mon smartphone quand je le souhaite.							
3	ET: Being unable to get the news (eg, happenings, weather, etc) on my smartphone would make me nervous. FT: Être incapable d'avoir accès aux nouvelles (eg, actualités, météo, etc.) sur mon smartphone me rendrait nerveux·se.							
4	ET: I would be annoyed if I could not use my smartphone and/or its capabilities when I wanted to do so. FT: Je serais contrarié·e si je ne pouvais pas utiliser mon smartphone et/ou ses fonctionnalités quand je le souhaite.							
5	ET: Running out of battery in my smartphone would scare me. FT: Ne plus avoir de batterie sur mon smartphone me ferait peur.							
6	ET: If I were to run out of credits or hit my monthly data limit, I would panic. FT: Si je me retrouvais sans crédits ou si j'atteignais ma limite de données mensuelle, je paniquerais.							
7	ET: If I did not have a data signal or could not connect to Wi-Fi, then I would constantly check to see if I had a signal or could find a Wi-Fi network. FT: Si je n'avais pas de signal réseau ou ne pouvais pas me connecter au Wi-Fi, je vérifierais constamment si j'ai du réseau ou si je peux trouver un réseau Wi-Fi.							
8	ET: If I could not use my smartphone, I would be afraid of getting stranded somewhere. FT: Si je ne pouvais pas utiliser mon smartphone, j'aurais peur de rester bloqué·e quelque part.							
9	ET: If I could not check my smartphone for a while, I would feel a desire to check it. FT: Si je ne pouvais pas consulter mon smartphone pendant un certain temps, j'aurais envie de le consulter.							
ET: If I did not have my smartphone with me, FT: Si je n'avais pas mon smartphone avec moi,								
10	ET: I would feel anxious because I could not instantly communicate with my family and/or friends. FT: Je me sentirais anxieux·se parce que je ne pourrais pas communiquer instantanément avec ma famille et/ou mes amis.							
11	ET: I would be worried because my family and/or friends could not reach me. FT: Je serais inquiet·e parce que ma famille et/ou mes amis ne pourraient pas me joindre.							
12	ET: I would feel nervous because I would not be able to receive text messages and calls. FT: Je me sentirais nerveux·se de ne pas pouvoir recevoir des messages ou des appels.							
13	ET: I would be anxious because I could not keep in touch with my family and/or friends. FT: Je me sentirais anxieux·se parce que je ne pourrais pas rester en contact avec ma famille et/ou mes amis.							
14	ET: I would be nervous because I could not know if someone had tried to get a hold of me. FT: Je serais nerveux·se parce que je ne pourrais pas savoir si quelqu'un a essayé de me contacter.							
15	ET: I would feel anxious because my constant connection to my family and friends would be broken. FT: Je me sentirais anxieux·se parce que ma connexion constante avec ma famille et mes amis serait rompue.							
16	ET: I would be nervous because I would be disconnected from my online identity. FT: Je serais nerveux·se parce que je serais déconnecté·e de mon identité en ligne.							
17	ET: I would be uncomfortable because I could not stay up-to-date with social media and online networks. FT: Je serais mal à l'aise parce que je ne pourrais pas être à jour avec les réseaux sociaux ou les réseaux en ligne.							

XSL•FO
RenderX

Item number	Item description	Response methods						
		1[a]	2	3	4	5	6	7[b]
18	ET: I would feel awkward because I could not check my notifications for updates from my connections and online networks. FT: Je me sentirais mal à l'aise parce que je ne pourrais pas vérifier mes notifications et réseaux sociaux.							
19	ET: I would feel anxious because I could not check my email messages. FT: Je me sentirais anxieux·se parce que je ne pourrais pas vérifier mes emails.							
20	ET: I would feel weird because I would not know what to do. FT: Je me sentirais mal à l'aise car je ne saurais pas quoi faire.							

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: strongly agree; French translation: fortement en accord.

[c]ET: English translation.

[d]FT: French translation.

XSL•FO
RenderX

Table 5. Smartphone Addiction Proneness Scale translation.

Item number	Item description	Response methods			
		1[a]	2[b]	3[c]	4[d]
1	ET[e]: My school grades dropped due to excessive smartphone use. FT[f]: Mes notes scolaires ont baissé à cause d'une utilisation excessive du smartphone.				
2	ET: I have a hard time doing what I have planned (study, do homework, or go to afterschool classes) due to using smartphone. FT: J'ai du mal à faire ce que j'avais prévu (étudier, faire mes devoirs, ou aller aux activités scolaires) à cause de l'utilisation du smartphone.				
3	ET: People frequently comment on my excessive smartphone use. FT: Les gens commentent fréquemment mon utilisation excessive du smartphone.				
4	ET: Family or friends complain that I use my smartphone too much. FT: Ma famille ou mes amis se plaignent que j'utilise trop mon smartphone.				
5	ET: My smartphone does not distract me from my studies (reversed item). FT: Mon smartphone ne me distrait pas de mes études. (énoncé inversé)				
6	ET: Using a smartphone is more enjoyable than spending time with family or friends. FT: Utiliser un smartphone est plus agréable que de passer du temps avec ma famille ou mes amis.				
7	ET: When I cannot use a smartphone, I feel like I have lost the entire world. FT: Quand je ne peux pas utiliser un smartphone, j'ai l'impression d'avoir perdu le monde entier.				
8	ET: It would be painful if I am not allowed to use a smartphone. FT: Ce serait douloureux si je n'étais pas autorisé·e à utiliser un smartphone.				
9	ET: I get restless and nervous when I am without a smartphone. FT: Je deviens agité·e et nerveux·se quand je suis sans smartphone.				
10	ET: I am not anxious even when I am without a smartphone (reversed item). FT: Je ne suis pas anxieux·se, même quand je suis sans smartphone. (énoncé inversé)				
11	ET: I panic when I cannot use my smartphone. FT: Je panique quand je ne peux pas utiliser mon smartphone.				
12	ET: I try cutting my smartphone usage time, but I fail. FT: J'essaie de réduire le temps d'utilisation de mon smartphone, mais j'échoue.				
13	ET: I can control my smartphone usage time (reversed item). FT: Je peux contrôler le temps d'utilisation de mon smartphone. (énoncé inversé)				
14	ET: Even when I think I should stop, I continue to use my smartphone too much. FT: Même lorsque je pense que je devrais arrêter, je continue de trop utiliser mon smartphone.				
15	ET: Spending a lot of time on my smartphone has become a habit. FT: Passer beaucoup de temps sur mon smartphone est devenu une habitude.				

[a]English translation: strongly disagree; French translation: fortement en désaccord.

[b]English translation: disagree; French translation: en désaccord.

[c]English translation: agree; French translation: en accord.

[d]English translation: strongly agree; French translation: fortement en accord.

[e]ET: English translation.

[f]FT: French translation.

Discussion

Principal Findings

The Problematic Use of Mobile Phones scale, the Smartphone Addiction Scale, the Smartphone Addiction Proneness Scale, and the Nomophobia Questionnaire for young adult students and adolescents have been developed and validated in English to measure problematic smartphone use, but they are not available in French. In this study, we provide their translations and linguistic validation in French.

Translating questionnaires into other languages can be complex, as terms in one language can be interpreted in different ways in another, leading to errors in meaning. For this reason, we

XSL•FO
RenderX

have followed the rigorous and standardized steps of linguistic validation defined by the forward/backward method [34].

The results of the translations showed that after 4 stages of translation, the content of the French questionnaires was identical to that of the English ones. The translators, adolescents, and young adults from different educational/professional and geographical backgrounds did not reveal any differences in understanding, suggesting that the cultural adaptation of the questionnaires was respected. These translations suggest that cross-cultural adaptation can be used for comparative purposes in other languages such as English, Portuguese, Arabic, or German.

However, 3 limitations need to be addressed in this study. First, the number of adolescents and young adult participants was less than 20. Although it would be more robust to interview more participants, the minimum number of participants for linguistic validation is not clearly defined [35]. Moreover, the adolescents and young adults provided only few comments on the clarity and understanding of the questionnaires. Second, although 4 volunteer translators were consulted, a more diversified panel of experts would have enabled us to avoid any cultural bias in the translation of the questionnaires. We were able to bring together 4 translators of different genders (ie, 2 men and 2 women) and from different disciplines (ie, eHealth, psychology, education, and economics) who did not have professional knowledge of the concept of problematic smartphone use. Despite these considerations, France is a country with a rich cultural diversity, which may include some regional and linguistic differences. We could have included native experts from different regions with marked linguistic peculiarities (eg, Brittany, Corsica). It is important to note that the 4 translators selected came from different regions of France, which ensures that we have reduced at least some of the cultural biases in translation. The last limitation was that this study focuses solely on linguistic validation without demonstrating quantitative validation. According to Terwee and colleagues [36], the translation process is the first necessary step for validation, but it is essential to verify content validity, internal consistency, criterion validity, construct validity, and reproducibility. This is particularly important here because problematic smartphone use, as measured by the questionnaires, is often assessed across multiple dimensions such as tolerance, withdrawal symptoms, overuse, or physical and psychological consequences [3]. These validation steps, especially regarding psychometric properties, are essential and must be completed before any use of these questionnaires in French. They can then be reliably used to better understand problematic smartphone use among French speakers and for conducting cross-cultural comparisons. These questionnaires will be relevant for gaining a better understanding of how problematic smartphone use is characterized among French adults and adolescents and can be utilized for screening and potentially monitoring patients with conditions associated with problematic smartphone use [5,37].

Conclusion

We succeeded in adapting and effectively translating 4 questionnaires assessing problematic smartphone use in a French-speaking context. This step remains a prerequisite to validating the questionnaires, which can serve as valuable research instruments for investigating and addressing issues related to problematic smartphone use in French-speaking countries and for making international comparisons. Given that problematic smartphone use is an emerging concern with significant health impacts [1], it is imperative that researchers and health care professionals use reliable measurement instruments.

Acknowledgments

This study would not have been possible without the support of Université Grenoble Alpes, Grenoble Alpes University Hospital Center, and LINKSIUM (Society of Technology Transfer and Startup, Grenoble Alpes). This work was partially funded by MIAI Cluster (ANR-23-IACL-0006) and the French National Research Agency (Agence Nationale de la Recherche) within the framework of the “Investissements d’avenir” program (ANR-10-AIRT-05, 2012; ANR-15-IDEX-02, 2016). The funders played no role in the study design, data collection, analysis and interpretation of data, or writing of this manuscript.

Authors' Contributions

IEB collected data, reviewed and assessed the results, wrote the original draft, wrote and reviewed the revised version of the manuscript, and acquired funding. MR collected analyzed and interpreted data, wrote the original draft, wrote and reviewed the revised version of the manuscript, and acquired funding. AB conceptualized and designed the study, reviewed and assessed the results, wrote and reviewed the revised version of the manuscript. NV conceptualized and designed the study, supervised the project, reviewed and assessed the results, wrote and reviewed the revised version of the manuscript and acquired funding. All authors reviewed and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Wacks Y, Weinstein AM. Excessive smartphone use is associated with health problems in adolescents and young adults. Front Psychiatry 2021;12:669042 [FREE Full text] [doi: 10.3389/fpsyt.2021.669042] [Medline: 34140904]

XSL•FO
RenderX

2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington DC: American Psychiatric Association Publishing; May 18, 2013.
3. Harris B, Regan T, Schueler J, Fields SA. Problematic mobile phone and smartphone use scales: a systematic review. Front Psychol 2020;11:672 [FREE Full text] [doi: 10.3389/fpsyg.2020.00672] [Medline: 32431636]
4. Notara V, Vagka E, Gnardellis C, Lagiou A. The emerging phenomenon of nomophobia in young adults: a systematic review study. Addict Health 2021 Apr;13(2):120-136 [FREE Full text] [doi: 10.22122/ahj.v13i2.309] [Medline: 34703533]
5. Ratan ZA, Parrish A, Zaman SB, Alotaibi MS, Hosseinzadeh H. Smartphone addiction and associated health outcomes in adult populations: a systematic review. Int J Environ Res Public Health 2021 Nov 22;18(22):12257 [FREE Full text] [doi: 10.3390/ijerph182212257] [Medline: 34832011]
6. Kwon M, Lee J, Won W, Park J, Min J, Hahn C, et al. Development and validation of a smartphone addiction scale (SAS). PLoS One 2013;8(2):e56936 [FREE Full text] [doi: 10.1371/journal.pone.0056936] [Medline: 23468893]
7. Yildirim C, Correia A. Exploring the dimensions of nomophobia: Development and validation of a self-reported questionnaire. Computers in Human Behavior 2015 Aug;49:130-137. [doi: 10.1016/j.chb.2015.02.059]
8. Merlo LJ, Stone AM, Bibbey A. Measuring problematic mobile phone use: development and preliminary psychometric properties of the PUMP scale. J Addict 2013;2013:912807 [FREE Full text] [doi: 10.1155/2013/912807] [Medline: 24826371]
9. Kim D, Lee Y, Lee J, Nam JK, Chung Y. Development of Korean smartphone addiction proneness scale for youth. PLoS One 2014;9(5):e97920 [FREE Full text] [doi: 10.1371/journal.pone.0097920] [Medline: 24848006]
10. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition. Washington DC: American Psychiatric Publishing; May 15, 1994.
11. Sfendla A, Laita M, Nejjar B, Souirti Z, Touhami AAO, Senhaji M. Reliability of the Arabic smartphone addiction scale and smartphone addiction scale-short version in two different Moroccan samples. Cyberpsychol Behav Soc Netw 2018 May;21(5):325-332. [doi: 10.1089/cyber.2017.0411] [Medline: 29762065]
12. Andrade ALM, Kim D, Scatena A, Enes CC, Enumo SRF, De Micheli D. Validity and reliability of the Brazilian version of the smartphone addiction scale-long version (SAS-LV). Trends in Psychol 2020 Dec 01;29(2):302-319. [doi: 10.1007/s43076-020-00046-y]
13. Ching SM, Yee A, Ramachandran V, Sazlly Lim SM, Wan Sulaiman WA, Foo YL, et al. Validation of a Malay version of the smartphone addiction scale among medical students in Malaysia. PLoS One 2015;10(10):e0139337 [FREE Full text] [doi: 10.1371/journal.pone.0139337] [Medline: 26431511]
14. Mokhtarinia H, Rafinia M, Shahbazi A, Khodaie Ardakani MR, Ghaedamini Harouni GR, Philip Gabel C. The cross-cultural adaptation of the smartphone addiction scale to Persian. Iran Rehabil J 2020 Mar 30:91-98. [doi: 10.32598/irj.18.1.919.1]
15. Lopez-Fernandez O. Short version of the smartphone addiction scale adapted to Spanish and French: Towards a cross-cultural research in problematic mobile phone use. Addict Behav 2017 Jan;64:275-280. [doi: 10.1016/j.addbeh.2015.11.013] [Medline: 26685805]
16. Jahrami H, Saif Z, Trabelsi K, Bragazzi NL, Vitiello MV. Internal consistency and structural validity of the nomophobia questionnaire (NMP-Q) and its translations: A systematic review with meta-analysis. Heliyon 2023 Apr;9(4):e15464 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e15464] [Medline: 37128316]
17. Gao Y, Dai H, Jia G, Liang C, Tong T, Zhang Z, et al. Translation of the Chinese version of the nomophobia questionnaire and its validation among college students: factor analysis. JMIR Mhealth Uhealth 2020 Mar 13;8(3):e13561 [FREE Full text] [doi: 10.2196/13561] [Medline: 32167480]
18. Adawi M, Bragazzi NL, Argumosa-Villar L, Boada-Grau J, Vigil-Colet A, Yildirim C, et al. Translation and validation of the nomophobia questionnaire in the Italian language: exploratory factor analysis. JMIR Mhealth Uhealth 2018 Jan 22;6(1):e24 [FREE Full text] [doi: 10.2196/mhealth.9186] [Medline: 29358161]
19. Graben K, Doering BK, Jeromin F, Barke A. Problematic mobile phone use: Validity and reliability of the Problematic Use of Mobile Phone (PUMP) Scale in a German sample. Addict Behav Rep 2020 Dec;12:100297 [FREE Full text] [doi: 10.1016/j.abrep.2020.100297] [Medline: 33364306]
20. Alshahwan H, Alosaimi F, Alyahya H, Mahyijari N, Shaik S. Arabic validation of problematic use of mobile phone scale among university students in Saudi Arabia. J Nat Sci Med 2020. [doi: 10.4103/jnsm.jnsm_9_19]
21. Barrault S, Durousseau F, Ballon N, Réveillère C, Brunault P. [Smartphone addiction: French validation of the internet addiction test-smartphone version (IAT-smartphone) and associated psychopathological features]. Encephale 2019 Feb;45(1):53-59. [doi: 10.1016/j.encep.2017.12.002] [Medline: 29397925]
22. Young KS. Internet addiction: the emergence of a new clinical disorder. CyberPsychology & Behavior 1998 Jan;1(3):237-244. [doi: 10.1089/cpb.1998.1.237]
23. Lopez-Fernandez O, Kuss DJ, Pontes HM, Griffiths MD, Dawes C, Justice LV, et al. Measurement invariance of the short version of the problematic mobile phone use questionnaire (PMPUQ-SV) across eight languages. Int J Environ Res Public Health 2018 Jun 08;15(6):1213 [FREE Full text] [doi: 10.3390/ijerph15061213] [Medline: 29890709]
24. Ting C, Chen Y. Smartphone addiction. In: Adolescent Addiction. England: Academic Press; 2020:215-240.
25. Chan SJ, Yeo KJ, Handayani L. Psychometric properties of the smartphone addiction proneness scale in a sample of Malaysian adolescents. IJPHS 2022 Sep 01;11(3):785. [doi: 10.11591/ijphs.v11i3.21971]

XSL•FO
RenderX

26. Randler C, Wolfgang L, Matt K, Demirhan E, Horzum MB, Beşoluk S. Smartphone addiction proneness in relation to sleep and morningness-eveningness in German adolescents. J Behav Addict 2016 Sep;5(3):465-473 [FREE Full text] [doi: 10.1556/2006.5.2016.056] [Medline: 27499228]
27. Zhang Q, Liu Z. Smartphone addiction proneness scale for adolescents? Chinese version: revision and evaluation on reliability and validity. Chinese Journal of Public Health 2021;37(7):1149-1152. [doi: 10.11847/zgggws1129783]
28. Andersson S, Granat L, Brännström M, Sandgren A. Translation, cultural adaptation, and content validation of the palliative care self-efficacy scale for use in the Swedish context. Int J Environ Res Public Health 2022 Jan 20;19(3):1143 [FREE Full text] [doi: 10.3390/ijerph19031143] [Medline: 35162163]
29. Yoon YE, Cho SY. Translation and linguistic validation of the Korean version of the Wisconsin Stone quality of life questionnaire. Int Neurourol J 2020 Mar;24(1):77-83 [FREE Full text] [doi: 10.5213/inj.1836238.119] [Medline: 32252189]
30. Poder TG, Carrier N, Mead H, Stevens KJ. Canadian French translation and linguistic validation of the child health utility 9D (CHU9D). Health Qual Life Outcomes 2018 Aug 29;16(1):168 [FREE Full text] [doi: 10.1186/s12955-018-0998-4] [Medline: 30157857]
31. Schettini A, Rillon P, Pirson G, De Coninck C. Completion of the French translation and linguistic validation of the BODY-Q. Psychiatr Danub 2020 Sep;32(Suppl 1):150-152 [FREE Full text] [Medline: 32890380]
32. Soulard J, Vaillant J, Baillet A, Vuillerme N. Translation and French linguistic validation of the bath ankylosing spondylitis functional index and the bath ankylosing spondylitis global score in patients with axial spondyloarthritis. Curr Rheumatol Rev 2023;19(4):449-454. [doi: 10.2174/1573397118666220829124234] [Medline: 36043728]
33. Translation and linguistic evaluation protocol and supporting material. WHODAS 2.0 Translation Package. URL: https://tinyurl.com/2etx7aea [accessed 2023-12-01]
34. Epstein J, Santo RM, Guillemin F. A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. J Clin Epidemiol 2015 Apr;68(4):435-441. [doi: 10.1016/j.jclinepi.2014.11.021] [Medline: 25698408]
35. Harachi TW, Choi Y, Abbott RD, Catalano RF, Bliesner SL. Examining equivalence of concepts and measures in diverse samples. Prev Sci 2006 Dec;7(4):359-368 [FREE Full text] [doi: 10.1007/s11121-006-0039-0] [Medline: 16845592]
36. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007 Jan;60(1):34-42. [doi: 10.1016/j.jclinepi.2006.03.012] [Medline: 17161752]
37. Yang J, Fu X, Liao X, Li Y. Association of problematic smartphone use with poor sleep quality, depression, and anxiety: A systematic review and meta-analysis. Psychiatry Res 2020 Feb;284:112686. [doi: 10.1016/j.psychres.2019.112686] [Medline: 31757638]

Abbreviations

DSM: Diagnostic and Statistical Manual of Mental Disorders

Edited by T de Azevedo Cardoso; submitted 25.10.23; peer-reviewed by B Imbert, H Li; comments to author 09.03.24; revised version received 26.06.24; accepted 23.09.24; published 26.02.25.

Please cite as:
El Boudi I, Riant M, Bellier A, Vuillerme N
French Versions of 4 English Questionnaires on Problematic Smartphone Use: Cross-Cultural Linguistic Translation and Adaptation Study
Interact J Med Res 2025;14:e53958
URL: https://www.i-jmr.org/2025/1/e53958
doi:10.2196/53958
PMID:

©Islam El Boudi, Mathilde Riant, Alexandre Bellier, Nicolas Vuillerme. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.02.2025. 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

Subtyping Service Receipt in Personality Disorder Services in South London: Observational Validation Study Using Latent Profile Analysis

Jack Steadman[1], PhD; Rob Saunders[2], PhD; Mark Freestone[1], PhD; Robert Stewart[3], MBBS, PhD

[1]Unit for Psychological Medicine, Wolfson Institute of Population Health, Barts and The London School of Medicine and Dentistry, Queen Mary, University of London, London, United Kingdom

[2]Centre for Outcomes Research and Effectiveness (CORE), Research Department of Clinical, Educational, and Health Psychology, University College London, London, United Kingdom

[3]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Jack Steadman, PhD
Unit for Psychological Medicine, Wolfson Institute of Population Health
Barts and The London School of Medicine and Dentistry
Queen Mary, University of London
Yvonne Carter Building
58 Turner Street
London, E1 2AB
United Kingdom
Phone: 44 7404042057
Email: jack.steadman1@nhs.net

Abstract

Background: Personality disorders (PDs) are typically associated with higher mental health service use; however, individual patterns of engagement among patients with complex needs are poorly understood.

Objective: The study aimed to identify subgroups of individuals based on patterns of service receipt in secondary mental health services and examine how routinely collected information is associated with these subgroups.

Methods: A sample of 3941 patients diagnosed with a personality disorder and receiving care from secondary services in South London was identified using health care records covering an 11-year period from 2007 to 2018. Basic demographic information, service use, and treatment data were included in the analysis. Service use measures included the number of contacts with clinical teams and instances of did-not-attend.

Results: Using a large sample of 3941 patients with a diagnosis of PD, latent profile analysis identified 2 subgroups characterized by low and high service receipt, denoted as profile 1 (n=2879, 73.05%) and profile 2 (n=1062, 26.95%), respectively. A 2-profile solution (P<.01) was preferred over a 3-profile solution, which was nonsignificant. In unconditional ($t_{3941,3939}$=19.53; P<.001; B=7.27; 95% CI 6.54-8) and conditional ($t_{3941,3937}$=−3.31; P<.001; B=−74.94; 95% CI −119.34 to −30.56) models, cluster membership was significantly related to receipt of nursing contacts, over and above other team contacts.

Conclusions: These results suggest that routinely collected data may be used to classify likely engagement subtypes among patients with complex needs. The algorithm identified factors associated with service use and has the potential to inform clinical decision-making to improve treatment for individuals with complex needs.

(Interact J Med Res 2025;14:e55348) doi:10.2196/55348

KEYWORDS

latent profile analysis; latent variable mixture modeling; service use; personality disorder; applied health research; electronic health records

XSL•FO
RenderX

Introduction

Personality disorders (PDs) are typically associated with poorer health outcomes [1] and high mental health service use [2], yet PD treatments remain poorly understood and are often not specific to the individual [3]. Understanding patterns of service use across the diagnostic spectra of PD comprises an important and novel area of research in the applied domain. There are also considerations of optimizing UK National Health Service (NHS) service capacity and improving relationships between services and service users, which are often challenging given the relational sequelae component within a PD diagnosis [4].

The longstanding needs of patients with PDs require effective management and treatment approaches. Being able to understand likely patterns of service use could help allocate resources in secondary care, given limited resources. Predictive modeling for more personalized treatments is vanishingly rare in conventional health care models of service delivery [5], yet there are growing calls in the methodology and policy literature for improvements in both treatments and relationships [6] for patients with complex needs.

Effective prediction of treatment outcomes is an area of continued interest in general mental health care [7]. To our knowledge, this body of work has yet to be applied to PD treatments; however, there are promising opportunities offered by large-scale applied health data research, greater availability of relevant datasets [8], and increasing sophistication of relevant computer software and analytical approaches. This paper expounds formative work in this sphere by conducting formal validation in understanding diagnostic patterns of service receipt, as a key step toward formal prediction and more tailored treatments.

Given the engagement difficulties that come with a PD diagnosis [9], there is clinical interest in better understanding patterns of service use across the swathe of *International Classification of Diseases, Tenth Revision* (*ICD-10*) PD diagnoses [10,11]. Enabling services to predict likely patterns of service contact can help personalize the treatment pathway, given likely service demand, and optimize resource allocation. In turn, if treatment pathways can be tailored to the needs of service users at presentation, this may reduce the propensity for harmful treatments [12], directly improve therapeutic relationships, indirectly reduce long-term service use [13], and aid capacity management and resource allocation [14].

Individuals with a diagnosis of PD are reported as misunderstood, at least in mental health care [15]. In addressing this impasse, this study provides mental health services with the necessary functionality to move away from uniform PD treatments and toward more personalized approaches [16]. Decision support studies in health care are gaining momentum [17], and although the adoption of such systems has been slower in mental health care, the past decade has seen a rise in pilot implementations [18,19]. There is also a need for more advanced methodologies beyond “first-generation” regression-based approaches, which face increasing challenges and scrutiny in the age of data scientific [20], such as mediation and structural equation models, both in PD management [21] and in clinical psychology [22,23].

There has been growing interest in person-centered approaches [24] driven by calls for methodological and analytical advancements [25], efforts to improve access [26,27], initiatives to democratize treatments [28], and a rising focus on the personalization of mental health care [16]. These, in turn, have resulted in developments in theory [29] and modeling [30,31]. For example, latent class analysis (LCA) comprises an increasingly researched, novel approach, especially relevant in unpicking potential sources of observed heterogeneity [32,33]. This paper integrates these research perspectives using latent profile analysis (LPA). LPA is an extension of LCA, statistically described as finite Gaussian mixture models [34].

Person-centered approaches, such as clustering methods, LCA, and LPA, provide a useful first step in dimension reduction [35], argued as a key focus in reducing heterogeneity encountered in PD treatments, research, and theory [36]. In the context of psychiatric comorbidity, the number of potential co-occurring factors that make up disorders can challenge traditional variable-centered approaches, which may struggle to model all potential interactions [29]. In contrast, clustering approaches comprise an important first step in the aim to personalize treatments and treatment pathways, such as with physical medicine in cancer treatments [37]. In addition to these clinical and methodological motivations, the search for subgroups in complex mental health treatment is further motivated by relevant policy calls in the treatment literature [38].

The aim of this study was to characterize the patterns of service receipt in a large sample of patients with a diagnosis of PD in an urban, secondary care setting. The research question considered how to best characterize service receipt, as part of the specialist PD treatment pathway, using routinely available treatment, diagnostic, and demographic data.

Methods

Overview

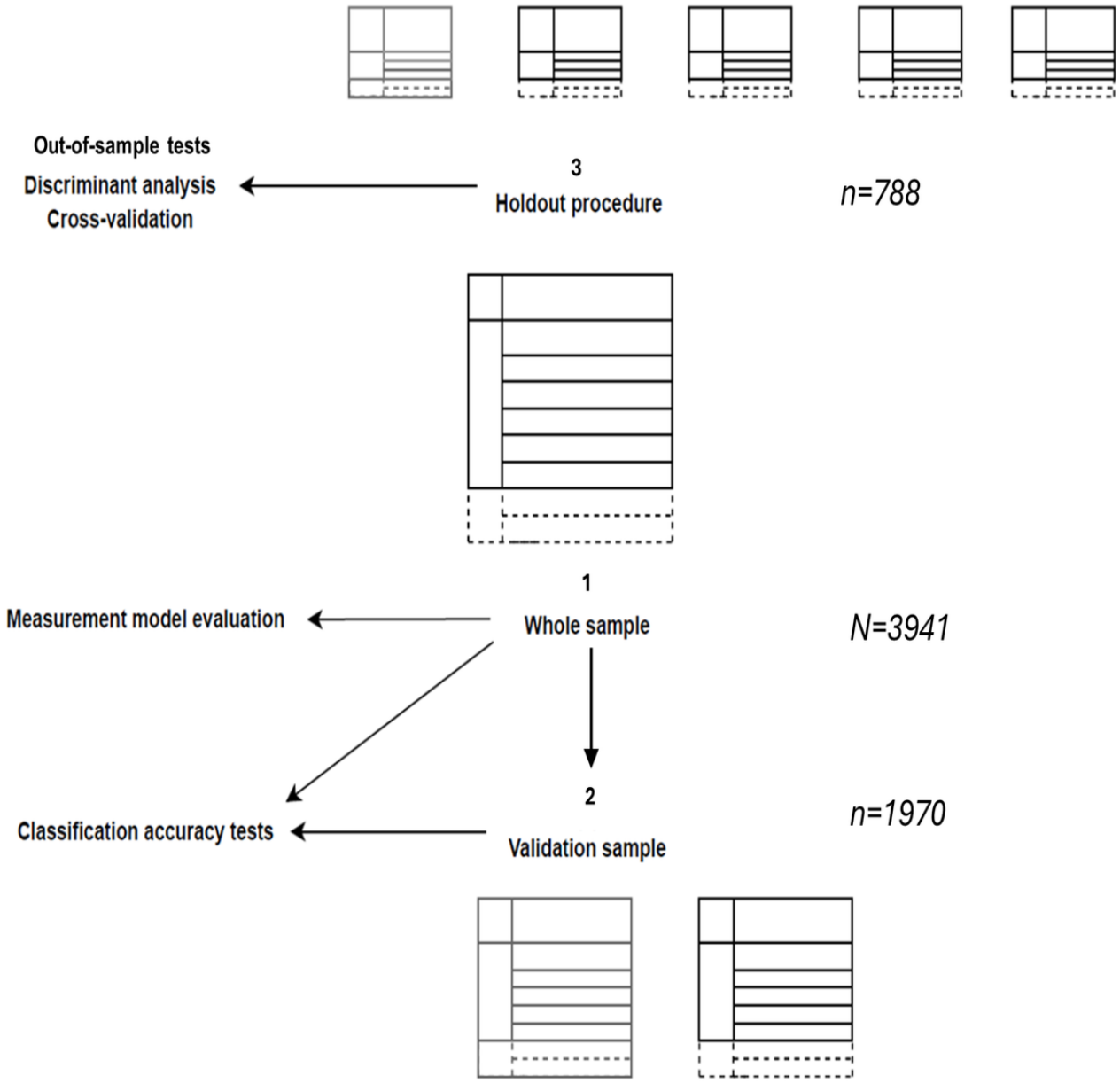
LPA was conducted to identify statistically distinct patterns of service contact within a sample of individuals with a diagnosis of PD receiving treatment in secondary care in South London. In addition to identifying the optimal class solution in this setting, formal validation aimed to inform generalization estimation, reporting both average-level and individual-level classification errors. This paper further extends LPA research through consideration and evaluation of the structural model and extends such evaluation through protocolized resampling.

Validation-focused analyses assessed the structural model [39], with parameters tested for relevance when considering cluster membership. Out-of-sample approaches to prediction are well researched in wider structural equation modeling perspectives, for example, partial least squares structural equation modeling [40]. Importantly, while findings of mixture models may prove interesting, classes may not necessarily reside in the wider population [41], further motivating explicit consideration of generalization in study design.

XSL•FO
RenderX

LPA guidelines suggest using 3 sampling frames as good practice to validate findings [42,43]. To begin, the initial whole sample was "sample frame 1," comprising the final analytical set. Validation samples randomly split the whole sample into two equal groups of 3941, resulting in "sample frame 2" [42,44] and 2 validation samples (1970/3941, 49.98%). Finally, cross-validation requires a holdout procedure [45], which partitions the dataset into test and training samples, to assess "out-of-sample" performance. This procedure split the whole sample into 5 parts, with 80% or $k-4$ parts designated as the training sample (3153/3941, 80%) and the remaining 20% or $k-1$ part designated as the test sample (788/3941, 20%). As shown in Figure 1, there were 3 sampling frames as a result.

Figure 1. Analysis plan and sampling procedure.



In this study, posterior probabilities of cluster membership were identified using sample frames 1 and 2. The comparison between sampling frames (Tables S1 and S3 in Multimedia Appendix 1) describes the formal relevance of the model regarding validation. Posterior probabilities of cluster membership were calculated from the e-step of the expectation maximization algorithm [46] and describe how the person i belongs to the class k [47,48], also described as their modal class assignment, according to the selected model [42].

Setting

Patient data for the study were extracted from the Clinical Record Interactive Search platform at the South London and Maudsley NHS Foundation Trust. South London and Maudsley (SLaM) provides comprehensive, near-monopoly specialist mental health services to a geographic catchment area of 4 boroughs of southeast London (Croydon, Lambeth, Lewisham, and Southwark) with a population of approximately 1.3 million residents.

XSL•FO
RenderX

Ethical Considerations

The Clinical Records Interactive Search (CRIS) system was set up in 2007-2008 through codevelopment of a data security and governance model with service users [49]. The CRIS system and related linked data have received successive research ethics approvals since 2008 as an anonymized data resource for secondary analyses (current approval: Oxford Research Ethics Committee C, approval number 23/SC/0257). The institutionally approved governance model has been described in detail elsewhere [50], but in summary, it includes automated deidentification of the record, standard researcher approval, a locally publicized opt-out facility, and an oversight committee with patient leadership and membership reviewing all data use.

Sample

The inclusion criteria comprised having an *ICD-10* PD diagnosis [41] at any time, being aged between 20 and 101 years at initial presentation, and being in contact with SLaM due to referral, although 10.38% (409/3941) of the sample had no recorded team contacts in the analysis set. Given the population of interest and generalization intention, those aged <20 years (318/3941, 8.07%) and >101 years (8/3941, 2.02%) were excluded to reduce possible inclusion bias and to encourage greater sample representativeness.

The inclusion criteria were applied over an exposure period of 11 years, with care episodes sampled between 2007 and 2018. This study sought to identify a matched, complete set of service use data at service users' first and final points of contact with services (time 1 and time 2, respectively). Time points were denoted by episode start and end dates, such that time 1 represented the start and end of the first episodes of care and time 2 represented the same at the final episodes of care. Dataset assembly resulted in a complete, matched set of data points, ascribed by time 1 and time 2 and organized by episode of care. An episode of care was calculated as the elapsed time between patients being referred to and in contact with services and subsequent discharge from secondary services. In the final matched pairs set, the average length of an episode of care was 131.86 (SD 309.69) days. Moreover, "discharge" is defined by SLaM Trust policy as "discharging a patient from an in-patient setting to a community team or from the Trust completely to primary care or another provider" [51]. Importantly, this follows when a patient is deemed to no longer benefit from either reduced risk or improved well-being [51] and suggests a shift in presentation and overall needs.

Working with the CRIS team, a custom data extract resulted in 82,265 episodes of care from 8510 service users. The inclusion of contacts from clinical teams data resulted in a reduced set of 6068 individuals, which was further processed to include complete "matched pairs" data for 3941 service users, at their first and final episodes of care (T1 and T2, respectively; Figure S1 in Multimedia Appendix 3). A complete dataset was preferred as there were no missing data, which strengthened the design using a large sample. T1 and T2 contained episode start and end dates, respectively, as discussed in the Statistical Analysis section. The authors acknowledge that the nature of this experimental design necessitated service users engaging with services on more than one occasion. Therefore, individuals with only a single episode of care (2127/3941, 53.97%) were omitted from the analysis. Notwithstanding, PD is typically a chronic condition [52,53], and in general, service users are likely to have ongoing needs and therefore present to services on more than one occasion.

Measurements

Quantitative Variables

Mixture modeling research typically uses cross-sectional data [54]. In the search for the best model, LPA compared model solutions using both repeated measures (T1 and T2) and T1 data only. Analytically, the episodic shaping of the data did not account for the timing or point of diagnosis, resulting in some patients being diagnosed after the initial T1 episode of care. Once again, however, given the longstanding nature of PD [52] and difficulties in accessing effective treatments [55], the precise timing of when a patient receives a diagnosis can be considered less important than it may initially appear.

Preliminary modeling [56] identified negligible "loading" or lambda values when including more granular detail regarding ethnicity (Table S1 in Multimedia Appendix 4). For example, including ethnicity categories such as Asian (λ=0.25) and mixed ethnicity (λ=−0.23) in the demographic construct failed to meet statistical significance, and the strongest performing model was the one that only included the presence or absence of White ethnicity.

Demographic data were modeled to include ethnicity (White or non-White), neighborhood-level deprivation, age, and gender (proportion of female and other). Deprivation was measured using the Multiple Indices of Deprivation [57]. This measure comprises nationally recorded data and measures multiple deprivation at the Lower Super Output Area level (a standard administrative geography encompassing approximately 1500 residents), including a nominal variable of 10 categories, with 1 describing most deprived and 10 describing least deprived [57].

Service receipt was modeled to include the total number of face-to-face contacts, as well as the number of unattended appointments, for each episode of care. Clinical input data were counted as the number of contacts from each treating team during each episode of care, that is, the cumulative total number of contacts within the specified time frame of a care episode. The number of contacts from care teams included medical, psychology, nursing, therapy, social work, and other. The CRIS system uses the General Architecture for Text Engineering [58], a natural language processing engine that captures recorded diagnostic statements. This output was used to ascertain PD diagnoses (Table S2 in Multimedia Appendix 4) in addition to prestructured primary diagnosis fields, which provide all options from the *ICD-10* [59].

ICD-10 lists a range of PD diagnoses, denoted by category membership, in contrast to *ICD-11*, which characterizes PD by severity. Accumulating evidence indicates limited support for the foregoing polythetic approaches [11,60]. To aid meaningful comparisons, diagnoses were collated and assigned to relevant clusters used routinely in psychiatric care (Table S2 in Multimedia Appendix 4). Diagnostic and Statistical Manual

XSL•FO
RenderX

cluster A diagnoses are characterized by "odd and eccentric" presentations, including paranoid, schizoid, and schizotypal PD diagnoses. These were distinguished from cluster B diagnoses, described as "overly dramatic, overly emotional, or unpredictable" presentations, as well as cluster C diagnoses, namely, "fearful or anxious" presentations.

As with ethnicity data, preliminary statistical modeling [56] highlighted the negligible impact on model performance of including more specific diagnostic information (Table S1 in Multimedia Appendix 2), beyond the presence or absence of cluster A diagnoses alone. Specifically, this was exemplified by low lambda loading values ($\lambda<0.40$) when including cluster B ($\lambda=0.25$), cluster C ($\lambda=-0.20$), and cluster other ($\lambda=-0.07$) onto the overall construct of diagnosis (refer to Tables S1 and S2 in Multimedia Appendix 4 for data cleaning descriptions of ethnicity and diagnostic data, respectively). Following these inferences, the proportion of *DSM* A diagnoses were taken as the proportion of *internalizing* diagnoses [61].

Statistical Methods

As formulated in the study by Nylund et al [62] and discussed in the study by Masyn [42], the null hypothesis in mixture modeling is the one-class model. Comparative likelihood ratio approaches, using the likelihood ratio test (LRT), describe and compare the likelihood of the two models under comparison, given the data [63]. In the first stage, different model solutions (namely, 2-, 3-, or 4-cluster solutions) were tested against this null sample. In the second stage, the number of groups, or "G," was varied using *mclust* [35], and relevant likelihood ratio statistics were then entered into *tidyLPA* [64]. Second-stage testing specified the null (ie, the 2-class solution) versus the alternative (ie, the 3-class solution) hypothesis. The LRT statistic then compares the likelihood of either model, given the data [63].

Model evaluation then proceeds from the measurement model, in terms of number of classes, to the structural model, in terms of the relation of parameters to classes. Across the range of "stepwise" configurations, 1- and 3-step procedures are typical of the literature [39,62], with different approaches varying in their inclusivity and efficiency [65,66]. In practice, comparing the relationship between parameters and clusters can assist in model evaluation. These relationships describe the structural model of the mixture model, and structural model evaluation compares these relationships.

Discriminant analyses are popular in machine learning perspectives on mixture models [67,68]. Discriminant validity assesses the classification accuracy of a mixture model by comparing the performance of a testing subsample on a training subsample (Figure 1). In this study, discriminant analysis was developed using the eigenvalue decomposition discriminant analysis [35,41]. The eigenvalue decomposition discriminant analysis constrains the discriminant analysis model to have a single component for each class, with the same covariance structure among classes. More recent discriminant analysis models are available [69]; however, these models impose stricter requirements, such as a different number of mixture components and covariance structures for each class, which were deemed unsuitable for this analysis.

Discriminant analysis was conducted over 10 folds. For each "fold," a separate dataset was assembled, comparing actual values versus those predicted by the model. As is typical in prediction studies, prediction error tests the predictive performance of the model on a holdout sample, with results summarized using common prediction metrics [45]. Typically, these include the root mean squared error, mean absolute error (MAE), and mean absolute percentage error (MAPE) [45,70]. In predictive applications, the MAE was selected as the metric of choice, given its greater resistance to distributional skewness [70]. To assist interpretation, both the MAE and MAPE are reported, with the MAPE calculated using the *Metrics* package [71]. The MAPE is calculated using the absolute error in each fold divided by the observed values that are evident for that fold. Averaging these fixed percentages yields the MAPE, which indicates how much error arises from the prediction model compared with the true value [72].

LPA was chosen for its flexibility in data requirements and its suitability for modeling a heterogeneous population that cannot be effectively represented by a single-component distribution [34]. Indicator variables included demographic variables, number of face-to-face contacts, number of did-not-attends, and number of contacts with clinical teams, per episode of care. Data preparation was assisted using the *tidyverse* package [73]. Analyses were conducted using the packages *mclust* [35] and *tidyLPA* [64] using R statistical software (R Core Team) [74,75].

Statistical Analysis

Fitting the Measurement Model

The purpose of this analysis was to identify the patterns of service receipt in a large sample of patients with a diagnosis of PD in a secondary care setting, with the goal of enhancing contemporary models of care [76]. LPA scholarship was further extended through formal validation, as a precursor to prediction, which used protocolized resampling to ascertain the discriminatory relevance and possible scale potential of the proposed model solution.

To establish the best-fitting model, a range of metrics were considered. These included the Vuong-Lo-Mendall-Rubin LRT [77] and the bootstrapped LRT [39,78]. Alongside likelihood ratio tests, fit statistics included the Akaike information criterion, the Bayesian information criterion (BIC), and values of entropy. Both the LRTs, the Vuong-Lo-Mendall-Rubin LRT and bootstrapped LRT compare the K model (the current model with a number of k profiles) to $K-1$ model (a model with one less profile). Significant P values from these tests assert that the K model fits the data better than a comparable model with one less profile [43]. Conversely, a nonsignificant P value ($P\leq.05$) indicates that the model with one less profile provides a better fit for the data, with more parsimonious models preferred. Smaller Akaike information criterion and BIC values indicate better model fit, while higher values of entropy suggest higher accuracy in classification of the model.

As no prior hypotheses for the number of profile groups were suggested, model selection compared solutions derived from either a repeated measures design, consisting of T1 and T2 data, or a contemporaneous design, consisting of T1 data only.

XSL•FO
RenderX

Episodes of care were defined across two levels: (1) at the patient level, including demographic and diagnostic information, and (2) at the episode level, including "start" and "end" date of episodes, whereby service receipt measures were collected as properties, or descriptors, of that episode.

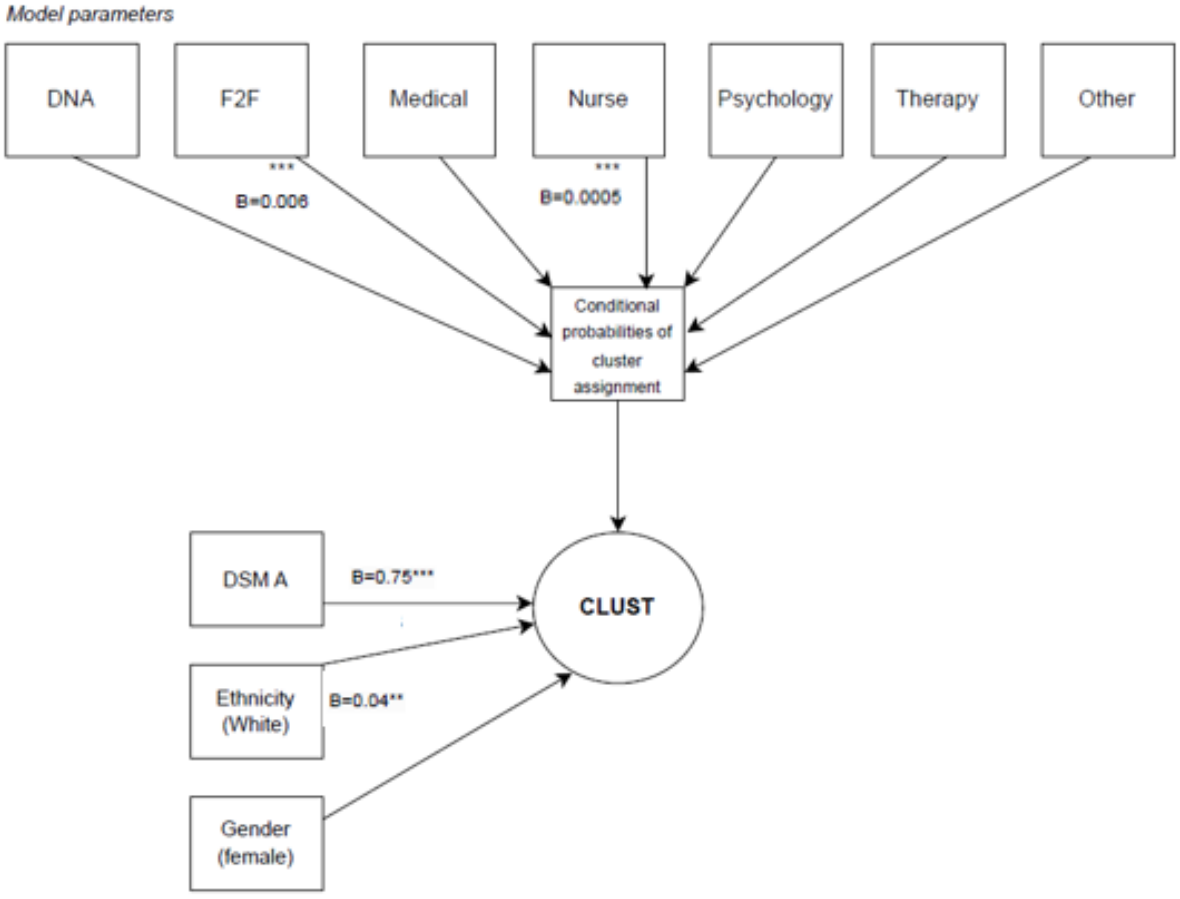
Episode descriptors were used as indicators for the LPA, including sociodemographic measures (age, ethnicity, and gender); diagnostic factors (percentage of internalizing diagnoses); general measures of contact (the number of face-to-face contacts and the number of did-not-attends); and the number of contacts from specific clinical teams (medical, nurse, psychology, therapy, and other).

Fitting the Structural Model

Assessment of the structural model considered the explanatory relevance of parameter values on cluster membership. Exogenous, pretreatment variables (ie, demographics) were limited to 1-step evaluation procedures. For service contact parameters, inferential tests were developed for both 1- and 3-step procedures [39]. Three-step procedures assert the importance of conditional probability of class membership in post hoc testing. Corrected by assignment probability [42], the 3-step procedure helps control for the individual validity of profiles. In contrast, the 1- and 2-step approaches can often omit the nonzero classification accuracy of derived clusters [39]. However, in the case of models with high classification accuracy, 1-step procedures may be justified [65,79]. As such, this paper reports both 1- and 3-step procedures, comparing unconditional versus conditional structural mixture models.

For the adapted 3-step procedure, the first step identified and enumerated cluster solutions, which were compared using model fit statistics. In the second step, conditional probabilities were extracted from the e-step of the EM algorithm using *mclust* [35] and vectorized "row wise" at the level of each individual. For the third step, individual conditional probabilities were included as covariates (ie, general linear models mapped the interaction between parameters with conditional probabilities). General linear models compared cluster assignment as a function of each parameter. Results of these analyses are summarized in Figure 2 and presented in Table 1.

Figure 2. Conditional structural model evaluation showing the significance of paths and pretreatment covariates in determining cluster membership.



XSL•FO
RenderX

Table 1. Direct and controlled effects of cluster membership and additional covariates on parameters.

	t test (*df*)	Direct effect, 95% CI	*t* test (*df*)	Controlled effect (95% CI)
Engagement				
Did-not-attend	19.05 (3941,3939)	4.30 (3.85 to 4.74)	0.05 (3941,3937)	2 (4 to 10)
Face to face	28.70 (3941,3939)	31.33 (29.19 to 33.47)	−1.04 (3941,3937)	−68.86 (−199.13 to 61.41)
Clinical contacts				
Medical	23 (3941,3939)	4.91 (4.49 to 5.33)	−1.30 (3941,3937)	−16.89 (−42.38 to 8.60)
Nurse	19.53 (3941,3939)	7.27 (6.54 to 8)	−3.31[a] (3941,3937)	−74.94 (−119.34 to −30.56)
Psychology	23.83 (3941,3939)	5.71 (5.24 to 6.18)	1.72 (3941,3937)	25.01 (−3.57 to 53.59)
Therapy	15.37 (3941,3939)	3.21 (2.80 to 3.61)	1.07 (3941,3937)	13.57 (−11.35 to 38.49)
Other	12.64 (3941,3939)	4.84 (4.09 to 5.59)	0.45 (3941,3937)	10.60 (−35.14 to 56.34)

[a]*P*<.001.

Latent Profile Cross-Validation Strategy

Cross-validation uses the data to test "the predictive power" of the model [80]. For the procedure, cross-validation splits the dataset into k subsamples of equal size, repeating k times, each time using a different test set (Figure 1). The best-fitting parameters of the model are then fit using k−1 subsamples, with the model's error then computed on the unused subsample or test set [81].

In line with the discriminant analysis procedure, cross-validation was conducted using sample frames 1 and 2 (in sample and out of sample, respectively), as shown in Figure 1. The Brier score [82,83] indicates the predictive usefulness of the model for individuals who, theoretically, may be unseen. Applied research in mental health has shown increased experimentation with the Brier score as a forecasting tool, such as for predicting suicidal ideation in individuals at near-term suicide risk [84] or identifying patients likely to drop out of psychotherapy [85]. In related psychotherapy research, the Brier score ranges from 0 (best prediction) to 1 (worst prediction), and it measures the accuracy of probabilistic predictions [85].

A detailed sampling strategy enabled both explanatory and predictive appraisal of the derived model solution, as shown in Figure 1. As well as the whole sample (frame 1), validation samples (frame 2) divided the dataset randomly, 2 × 50%. LPA was conducted on both frames 1 and 2 (total sample and validation samples, respectively), as shown in Figure 1, informing sample comparisons and confirming the profile structure [43]. Subsequently, LPA was rerun on the whole sample to aid interpretation of the findings.

Sample frame 3 for discriminant analyses split the total sample into 5 parts, with 80% (3153/3941) or *k*−4 parts designated as the training sample and the remaining 20% (788/3941) or *k*−1 part designated as the test sample. The holdout procedure was repeated over 10 folds, in line with best practices [45]. Discriminant analyses tests were then applied over sampling frames 1 and 3 (Figure 1), with cross-validation describing classification accuracy for both total and holdout sampling frames using the classification error and Brier scores. As discussed earlier, the Brier score is a proper score function that measures the accuracy of probabilistic predictions [85]. More technically, the Brier score comprises the mean square difference between the true classes and predicted probabilities [82,83].

Latent Profile Resampling Testing Strategy

Discussions on prediction-focused structural equation modeling consider discriminant validity in terms of in-sample versus out-of-sample perspectives [45]. Bootstrapping has gained prominence in sample validation; however, this remains a decidedly in-sample approach. In this study, larger sample sizes permitted a combination of k-fold assessments of discriminant validity, combined with cross-validation modeling, which ascertain both in-sample and out-of-sample classification accuracy.

LPA was then conducted on the "T1 and T2" dataset, while a holdout procedure (Figure 1) tested the predictive relevance of the profile solution across 10 partitions, or "folds," of this procedure. Predictive performance was summarized by prediction metrics, namely, the MAE and MAPE [86,87]. The MAPE was calculated using the absolute error in each fold divided by the observed values that are evident for that fold. Averaging these fixed percentages yields the MAPE, which indicates how much error arises from the prediction model compared with the true value [72].

Results

This study broadly follows the reporting guidelines from the Consolidated Standards of Reporting Trials (CONSORT) [88], although it specifically adheres to the Strengthening the Reporting of Observational Studies in Epidemiology [89] and Reporting of Studies Conducted Using Observational Routinely-Collected Data [90] guidelines, which were more suited for improving the reporting of observational epidemiological research.

Participants

The flow of participants throughout the study is shown in Figure S1 in Multimedia Appendix 3. In brief, the initial analytical set comprised 82,265 episodes and 8510 patients. Data were prepared to include team contacts and ensure completeness,

XSL•FO
RenderX

resulting in 170,365 episodes and 6421 patients. Further processing resulted in the final analytical set of 7882 episodes from 3941 patients to optimize the analysis, with this complete dataset having no missing values.

Descriptive Statistics

Tables 2 and 3 display the distribution of patient characteristics from the LPA and across sampling frames. Descriptive statistics for holdout samples were limited to the initial fold (of 10). Comparing latent profiles to the whole sample means and distribution provides an understanding of the characteristics of each group of patients across identified latent profiles (LPs)—LP1 and LP2—in the 2-profile solution. Service receipt, measured as the number of contacts from clinical teams, was markedly higher for LP2 than for LP1. The LPA best characterized service receipt as being either high or low, captured through LP2 and LP1, respectively. While LP1 counted no psychology contacts, LP2 on average had almost 6 per episode of care. This pattern was consistent for medical teams and, in particular, nursing teams.

In general, LP2 received treatment at a rate 10 times higher than LP1. The derived profiles were not differentiated by sociodemographic characteristics; both latent profiles had similar proportions of female and White individuals as well as comparable levels of neighborhood deprivation. In total, 11.3% (120/1062) of LP2 had internalizing diagnoses, with no such diagnoses present in LP1.

Table 2. Latent profiles and associated patient characteristics for the full sample; indicator values for sample frames 2 and 3.

	Sample frame 1 (N=3941)	LP1[a] (n=2879; 73.05%)	LP2 (n=1062; 26.95%)	Sample frame 2		Sample frame 3 (holdout)	
				Validation sample a (n=1970)	Validation sample b (n=1970)	Train (n=3153)	Test (n=788)
Demographics							
Age (y), mean (SD)	44.33 (14.30)	44.10 (14.1)	44.95 (14.90)	43.90 (14.23)	44.38 (14.29)	44.49 (14.26)	43.68 (14.41)
Gender (n=2482, 63%)							
Female, mean (SD)	2482.8 (1891.7)	1805.1 (1381.9)	660.6 (1062)	1221.4 (965.3)	1241.1 (945.6)	1986.4 (1513.4)	472.8 (386.1)
Deprivation							
IMD[b], mean (SD)	3.68 (1.73)	3.66 (1.75)	3.70 (1.67)	3.74 (1.76)	3.64 (1.65)	3.66 (1.69)	3.74 (1.87)
Ethnicity (n=1261, 32%)							
White, mean (SD)	1261.1 (1852.3)	878.1 (1324.3)	386.6 (509.8)	650.1 (925.9)	650.1 (925.9)	1008.9 (1481.9)	252.16 (370.4)
Diagnostic data							
Internalizing (y), mean (SD)	118.23 (669.9)	0 (0)	120 (339.8)	59.1 (315.2)	59.1 (354.6)	94.59 (536)	23.64 (126.08)

[a]LP: latent profile.

[b]IMD: indices of multiple deprivation.

XSL•FO
RenderX

Table 3. Latent profiles (LPs) and associated engagement characteristics across sampling frames; indicator values for sample frames 2 and 3.

	Sample frame 1 (N=3941)	LP1 (n=2879; 73.05%)	LP2 (n=1062; 26.95%)	Sample frame 2		Sample frame 3 (holdout)	
				Validation sample a (n=1970)	Validation sample b (n=1970)	Train (n=3153)	Test (n=788)
Engagement, mean (SD)							
F2F[a]	12.88 (33.44)	4.43 (5.17)	35.78 (58)	12.06 (30.25)	13.03 (33.93)	12.93 (33.49)	12.65 (33.24)
Did-not-attend	1.59 (6.57)	0.43 (1.03)	4.73 (12)	1.57 (6.51)	1.59 (6.79)	1.60 (6.62)	1.55 (6.35)
Clinical team input, mean (SD)							
Medical	2.32 (6.33)	1 (1.50)	5.90 (11.2)	2.20 (6.13)	2.33 (6.78)	2.32 (6.17)	2.34 (6.94)
Nurse	4.11 (10.86)	2.15 (3.36)	9.43 (19.20)	3.98 (10.23)	4.14 (10.13)	4.13 (11.41)	4.05 (8.29)
Psychology	1.54 (7.13)	0 (0)	5.70 (12.90)	1.49 (7.01)	1.46 (6.76)	1.64 (7.57)	1.15 (4.98)
Therapy	0.91 (5.98)	.05 (0.21)	3.25 (11.20)	0.74 (3.79)	0.94 (6.52)	0.89 (5.36)	0.99 (8.01)
Other	1.56 (10.88)	0.25 (0.72)	5.10 (20.5)	1.23 (9.27)	3.64 (1.65)	1.55 (11.09)	1.60 (9.99)

[a]F2F: face to face.

Evaluation of the Measurement Model

Analyses identified 2 profiles emerging from the contemporaneous (T1) dataset and 4 profiles from the repeated measures dataset (T1 and T2; Table 4). Profiles were enumerated according to the BIC using mclust [35] for T1 versus T1/T2 dataset, selecting 2 and 4 profiles, respectively. The comparison of classification and likelihood indices shown in Table 4 suggested that the 2-profile solution from T1 data was preferred, showing superior model fit. Stage-1 LRT tests compared 2 and 3 class solutions against the null hypothesis, that is, the 1-class model. Both solutions demonstrated superior likelihood than the null hypothesis; however, the 2-profile solution was preferred. Stage-2 LRT tests compared 3-profile (the alternative) versus 2-profile (the null) solutions and failed to demonstrate any advantage of a 3-profile solution over a 2-profile solution.

Table 4. Model fitting statistics[a].

Fit values	1-cluster	2-cluster	3-cluster	4-cluster
Log-likelihood criteria				
Log likelihood	−119,761	−72,497.47	−81,918.54	−118,721.9
BIC[b]	−240,267.2	−146,402.4	−164,267.6	−245,317.3
Classification criteria				
Entropy	—[c]	13.14	90	29
Normalized entropy difference	—	0.034	0.025	0.0052
LRTs [d]	1 versus 2	2 versus 4	3 versus 1	3 versus 2
VLMR-LRT[e]	90,868.55[f]	−90,624.52	92,661.71[f]	ns[g]
B-LRT[h]	94,527.1[i]	—	—	—

[a]Multivariate indicators: multivariate skewness=1425.64 (statistic 936,405.6); multivariate kurtosis=2840.37 (statistic 4576.15). For a formal mathematical definition, see Figure S2 in Multimedia Appendix 3.

[b]BIC: Bayesian information criterion.

[c]Not applicable.

[d]LRT: likelihood ratio test.

[e]VLMR-LRT: Ven-Lo-Mendall-Rubin likelihood ratio test.

[f]*P*<.001.

[g]ns: nonsignificant.

[h]B-LRT: bootstrapped likelihood ratio test.

[i]*P*<.01.

In model selection, comparative LRTs described how a 3-cluster solution significantly outperformed the 1-class model but not the 2-class model (Table 4); therefore, a 2-cluster solution was preferred.

XSL•FO
RenderX

Posterior probabilities were calculated as the proportion of correctly assigned patients to the respective clusters. These probabilities are "posterior," in the sense that they are calculated *after* individuals have been assigned a cluster membership [91]. Table S2 in Multimedia Appendix 1 displays classification indicators and posterior probabilities, and the classification error of the model was also negligible, being near 0. The Brier scores further demonstrate near-perfect classification accuracy of the model in forecasting individuals' likely engagement subtype, given the values of routinely collected indicators. High posterior probabilities are also considered clear indicators of classification accuracy, with values ideally exceeding 0.80, in line with established guidelines [66].

Evaluation of the Structural Model

Findings from evaluation of the measurement model, taken with the near-perfect degree of classification accuracy (Table S3 in Multimedia Appendix 1), suggest the usefulness of combined parameters in predicting cluster membership.

Relating parameter values to clusters is expressed in the language of direct and controlled direct effects [92]. The direct effect describes the (unconditional) relationship between parameters and clusters, while the controlled direct effect describes this relationship based on the individual probabilities of cluster assignment. In the unconditional model, all parameters determined class assignment, apart from gender (Table 5). Figure 2 displays conditional model evaluation, where face-to-face contacts ($t_{3940,3939}$=5.73; P<.001; B=0.006; 95% CI 0.005-0.006) and the number of nursing contacts ($t_{3940,3939}$=5.52; P<.001; B=0.0005; 95% CI −0.003 to 0.0006) significantly determined class assignment.

Likelihood-based approaches tend to dominate the mixture modeling landscape; however, there can be inferential challenges when adhering to the orthodoxy [63,93]. A more Bayesian approach may consider the likely parameter value, given the cluster membership. Structural model evaluation using this approach followed the same 3-step procedure as advised by Nylund-Gibson et al [39] and described in the section Fitting the Structural Model. Results from these analyses are presented in Table 1 and further detailed in Figure 3.

In the unconditional model and of all care contacts, psychology demonstrated the strongest relationship to clusters ($t_{3940,3937}$=23.83; P<.001; B=0.02; 95% CI 0.02-0.02); however, this was not upheld in the conditional model ($t_{3940,3937}$=1.72; P=.09). Overall, key prognostic targets of service receipt given the knowledge of cluster membership were nurse contacts. In both unconditional (t_{3940}=19.53; P<.001; B=7.27; 95% CI 6.54-8) and conditional (t_{3939}=5.52; P<.001; B=−74.94; 95% CI −119.34 to −30.56) models, cluster membership was significantly related to the receipt of nursing contacts.

Further testing modeled an interaction term between nursing contacts and cluster membership in determining nonattendance, as shown in Figure 3. The interaction term was significant both for face-to-face contacts ($t_{3940,3937}$=5.81; P<.001; B=0.74; 95% CI 0.49-0.99) and did-not-attends ($t_{3940,3937}$=3.62; P<.001; B=0.12; 95% CI 0.06-0.20), with confirmatory testing indicating that the rate of did-not-attends did not significantly differ from a Poisson distribution (N=3941, χ^2_8=10.3; P=.20).

Table 5. Direct and controlled effects of parameters and additional covariates on cluster membership.

	t test (*df*)	Direct effect, 95% CI	*t* test (*df*=3939)	Controlled effect, 95% CI
Engagement				
Did-not-attend	2.93 (3940)	0.004 (0.001 to 0.006)	1.53	0.0004 (−0.0001 to 9.86)
Face to face	20.76 (3940)	0.005 (0.004 to 0.006)	5.73[a]	0.006 (0.005 to 0.006)
Clinical contacts				
Medical	23 (3940)	0.024 (0.022 to 0.026)	1.38	0.0003 (−0.0001 to 0.0007)
Nurse	19.53 (3940)	0.01 (0.01 to 0.01)	5.52[a]	0.0005 (−0.003 to 0.0006)
Psychology	23.83 (3940)	0.02 (0.02 to 0.02)	1.72	0.0003 (−0.001 to 0.003)
Therapy	15.37 (3940)	0.02 (0.02 to 0.02)	0.15	0.0002 (−0.002 to 0.003)
Other: sociodemographics				
Internalizing diagnoses	19.07 (3940)	0.75 (0.68 to 0.81)	—[b]	—
Ethnicity (White)	2.97 (3940)	0.04 (0.01 to 0.06)	—	—
Gender (female)	1.16 (3940)	0.001 (−0.01 to 0.04)	—	—

[a]P<.001.

[b]Not applicable.

XSL•FO
RenderX

Figure 3. Plot showing the interaction term between nursing contacts and cluster in determining the rate of did-not-attends (DNAs).

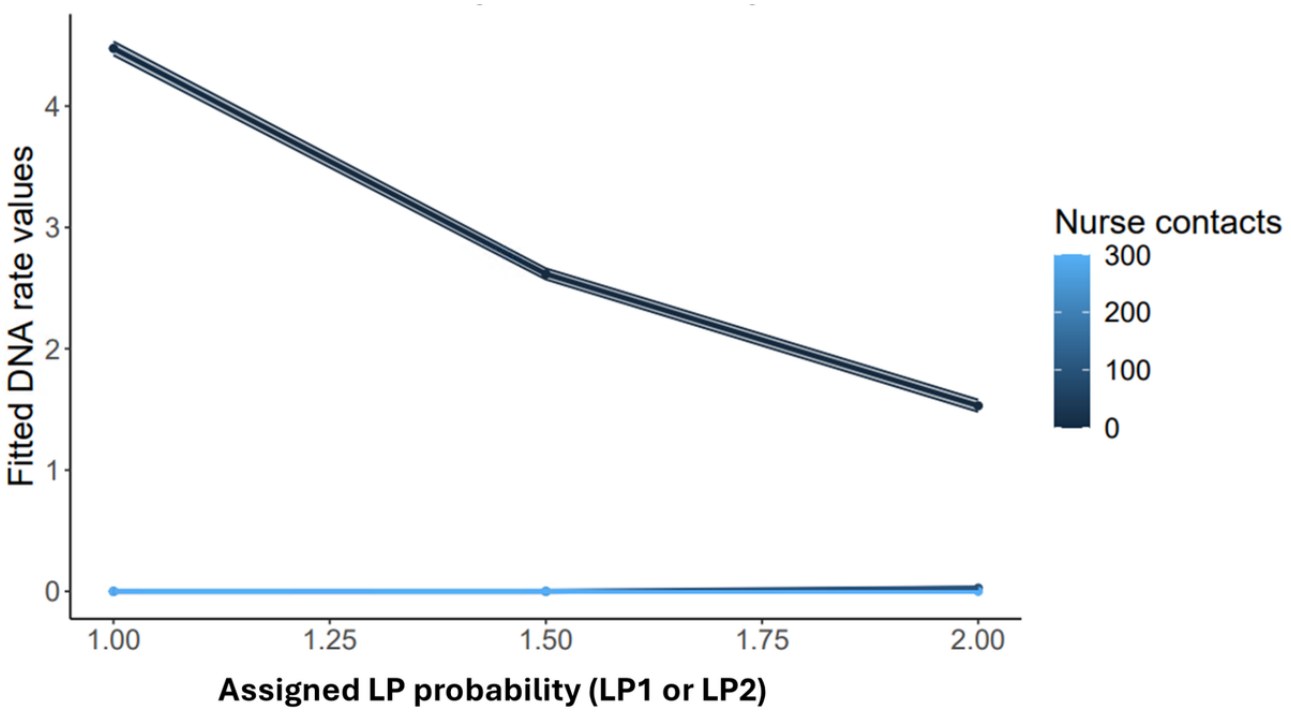


Figure 3 shows this relationship visually and indicates how did-not-attend rates vary between clusters as a function of nursing contacts. The assignment of profiles is shown probabilistically between either latent profile 1 and latent profile 2, which allows for assignment likelihood or uncertainty in clinical practice. For LP2, the number of nursing contacts made less of a difference in fitted rates of did-not-attends, compared to LP1. As it represented the majority of the sample, LP1 was more sensitive to the effect of nursing contacts regarding fitted rates of attended appointments.

Resampling Performance

Discriminant analysis modeling indicated high classification accuracy of the selected model (Table S3 in Multimedia Appendix 1). Both the MAE and MAPE were taken as summative indicators regarding the overall prediction error of the model. Over 10 folds of the holdout procedure, or train × test runs, the "grand" MAE was 0.004 (Table S3 in Multimedia Appendix 1), meaning that the model predicted correct class assignment with near-perfect accuracy. Conversely, out-of-sample tests on the replicative utility of engagement measures were less favorable. For face-to-face contacts, the MAE was 11.27, whereas for the number of did-not-attends, the MAE was 1.97. The MAPE may aid interpretation of these results, specifying that the number of face-to-face contacts erroneously predicts cluster membership in 11% (86/788) of cases (of 10 "folds"; Table S2 in Multimedia Appendix 1).

In the explanation-focused in-sample approach, the Brier score was 0.005. From an out-of-sample perspective, the average Brier score was 0.19 (Table S1 in Multimedia Appendix 1). The average classification error over 10 folds was 0.005, indicating substantially high classification accuracy. In addition, the average (out-of-sample) Brier score was 0.19 (SE 0.005), suggesting a nearly 80% rate of forecasted classification accuracy from an out-of-sample perspective.

Discussion

Summary of Findings

This study aimed to characterize the patterns of service receipt in a large sample of patients with a diagnosis of PD in an urban, secondary care setting. The main findings of this study were identification of 2 LPs that best characterize service receipt in specialist PD treatments. LP2, at 26% (1062/3941) of the sample, received almost 5-fold as many nursing contacts as LP1 (9.43 vs 3.98) and 6-fold as many medical contacts (5.90 vs 1). In addition, general service use was almost 10 times greater in LP2 than LP1. Sociodemographic characteristics were not associated with LPs, but diagnostic factors were associated. Specifically, LP1 had no internalizing diagnoses. These findings suggest important distinctions of service receipt as a function of routinely collected data and suggest the discriminatory relevance of such descriptors in determining likely profile membership.

LPA identified statistically reliable categories of service use using routine observational data for those with a diagnosis of PD. Profiles distinguish service receipt for those with a diagnosis of PD and were replicated across a range of sampling frames. Discriminant analyses and cross-validation modeling suggest substantial discriminatory performance in secondary settings using these types of data. This study described 2 LPs, primarily distinguished by diagnostic information. Of the approximately 3000 patients in LP1 (n=2879, 73.05%), none had internalizing PD diagnoses. In inferential testing, patients in LP1 were twice as likely to have less service contact.

Interpretation

The literature discussing PD and service use describes greater service use in the general presence of PD [94,95]; however, this has yet to be robustly applied between different PD diagnoses.

XSL•FO
RenderX

One study delineates the interplay between personality (disorder) and health [96], and there is rigorous work, such as comparisons of service use based on ethnicity in PD treatments [97]. Despite economies of scale offered by population-based interventions, these programs should be customized for subgroups on the basis of personality [96]. This study details foundational work in this area using a large sample in secondary care and attempts to export this comparative focus to diagnostic factors within PD treatments.

In a large study using administrative data in Quebec, emergency department visits were compared between "externalizing" or *Diagnostic and Statistical Manual* cluster B PD and schizophrenia, those diagnosed with both conditions, and the general population [98]. General practitioner visits were highest in the cluster B group, while the use of psychiatric (vs nonpsychiatric specialist or generalist) services differed in the presence of a PD diagnosis, with 1 mean annual visit versus 4, for schizophrenia.

Similar findings were captured in this study, which demonstrates lower service use for noninternalizing diagnostic groups. Greater service use was shown for the smaller latent profile, LP2. In practice, greater patterns of service use from the minority group arguably have predictive relevance in profile determination and tailoring treatment pathways accordingly. "Hotspotting," as it is known in general practice [99], is a data-driven practice that aims to identify patterns of "extreme" use [100]. Complex needs and PD populations are similarly characterized by high use [101]. This responsive concept is well known in general practice but has yet to be specifically applied to mental health care. In the case of PD and complex needs, general practice is often a key contact [102]. As demonstrated in this study, understanding individual (stratified) patterns of service use offers an example for expanding hotspotting practices, commonly used by general practitioners, to address historical challenges in generalist service provision for complex needs. By applying advanced techniques such as LPA, this approach can help create more tailored treatment pathways.

Therefore, this study argues that more data-enabled, proactive service provision, engendered by hotspotting practices, may assist modernization of mental health provision and encourage more individually intelligible and tailored treatment pathways. The assertion is that the determination of cluster membership, given the routinely measured parameters, may help inform similar practices in secondary care with complex needs. Related work in improving access to psychological therapies highlights 8 LPs [43], which contrasts with the solution presented in this study.

Interestingly, the rates of nonattendance were stable across profiles, averaging around 1 in 10, which is broadly in line with the literature in general practice [103]. Service use was poorly distinguished as a function of the treating team. Structural evaluation revealed nursing contacts and did-not-attends of key interest; modeling the interaction term showed that the fitted rate of did-not-attends differs between clusters as a function of nursing contacts. The implication is that more intensive treatment pathways with more nursing contacts result in greater fitted nonattendance for LP1. However, for LP2, greater nursing contacts had much less of an effect on did-not-attends.

The use of a large sample, the multiplicity of analyses, and the development of a protocolized resampling scheme for validation comprise key strengths of this study. Discriminant validity and classification accuracy tests showed near-perfect classification for individuals with a diagnosis of PD into LPs (Table S2 in Multimedia Appendix 1). Generalizability was explicitly considered pursuant to resampling and validation, which may have implications regarding wider (national) policy implementation. From an out-of-sample perspective, for example, the model accurately predicted cluster assignment in nearly 80% of cases, across 10 "folds" of the holdout procedure, or train × test runs, compared with a Brier score of 0.005 for in-sample model performance (Table S1 in Multimedia Appendix 1). At least, for explanatory purposes, the model was able to classify all cases correctly.

Limitations

A limitation of this study is that both exposure and outcome were measured over the same periods. This contemporaneous design was indicated in model selection; however, more frequent measurement may have resulted in more precise model solutions. Comparative modeling between both T1 and T1/T2 data suggested that a 2-cluster solution using T1 measurement showed superior model performance. In addition, the analytical design only included those who had more than 1 episode of care, which may exclude more sporadic or crisis presentations. Covariates were also drawn over time, rather than at presentation, rendering this study more suitable for validation than for strict prediction.

In addition, clinical outcomes were not assessed in this study, although there is scope to include these in future research. Finally, the direction of causality remains unclear—specifically, whether service receipt determines latent profile, or vice versa—along with the relationship between service receipt, profile, and clinical outcomes.

Conclusions

Truly "predictive" models of care are currently uncommon in mental health [5]. However, this study plays an important role in moving toward more predictive models of care in the medium term, following the formal validation of LPs and deployment of a rigorous resampling procedure in testing classification uncertainty at both average and individual levels. Through the identification and description of LPs, services are envisioned to be equipped to include such information in more personalized treatment approaches. More reflexively, such approaches can be seen as facilitating preventative models of care rather than reactive ones [104,105], representing a radical and, if not timely, necessary shift in care delivery.

XSL•FO
RenderX

Acknowledgments

The first author would like to thank Dr William Marsh for their assistance in the development of the required code to sample episodes of care as well as their supervisory assistance for work related to this study. In addition, JF would like to thank MF for their indispensable assistance and supervision throughout this study and for highlighting the funding opportunity. Finally, R Saunders helped inspire the conceptualization of this research, offering invaluable methodological and theoretical assistance at key periods, for which the authors are profoundly grateful. The authors would also like to thank the dedicated editors and peer reviewers from *JMIR Interactive Journal of Medical Research* for their insightful comments and attention to detail, which helped this study fulfill its potential. This report is an independent research funded by the National Institute for Health and Care Research Applied Research Collaboration North Thames. This paper represents an independent research partly funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, NIHR, or Department of Health and Social Care. This project was made possible through additional NIHR funding through the NIHR Short Placement for Applied Research Collaboration award.

R Stewart is partly funded by (1) the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London; (2) the NIHR Applied Research Collaboration South London at King's College Hospital NHS Foundation Trust; (3) the Datamind Health Data Research United Kingdom Mental Health Data Hub (Medical Research Council grant MR/W014386); and (4) the United Kingdom Prevention Research Partnership (Violence, Health and Society; MR-VO49879/1), an initiative funded by United Kingdom Research and Innovation Councils, the Department of Health and Social Care (England) and the United Kingdom–devolved administrations, and leading health research charities.

Data Availability

The datasets generated or analyzed during this study are not publicly available.

Authors' Contributions

The project was developed by JS with assistance from R Stewart and supervisory support from MF. Data were curated by R Stewart. JS analyzed the data and drafted the paper as first author. Formal analyses were conducted by JS. R Saunders assisted in conceptualization at key points of designing the methods and analysis. Resources were provided by R Stewart. All authors reviewed the manuscript, which was drafted by the first author, JS.

Conflicts of Interest

R Stewart received research support from GSK and Takeda in the last 3 years prior to manuscript submission. All other authors declare no conflicts of interest.

Multimedia Appendix 1
Additional validation and classification indices.
[DOCX File , 21 KB - ijmr_v14i1e55348_app1.docx]

Multimedia Appendix 2
Ethnicity and diagnostic data of total sample.
[DOCX File , 17 KB - ijmr_v14i1e55348_app2.docx]

Multimedia Appendix 3
Additional visual displays.
[DOCX File , 173 KB - ijmr_v14i1e55348_app3.docx]

Multimedia Appendix 4
Data cleaning procedures for categorical data.
[DOCX File , 18 KB - ijmr_v14i1e55348_app4.docx]

References

1. Moran P, Romaniuk H, Coffey C, Chanen A, Degenhardt L, Borschmann R, et al. The influence of personality disorder on the future mental health and social adjustment of young adults: a population-based, longitudinal cohort study. Lancet Psychiatry 2016 Jul;3(7):636-645 [FREE Full text] [doi: 10.1016/S2215-0366(16)30029-3] [Medline: 27342692]
2. Coid J, Yang M, Tyrer P, Roberts A, Ullrich S. Prevalence and correlates of personality disorder in Great Britain. Br J Psychiatry 2006 May;188:423-431. [doi: 10.1192/bjp.188.5.423] [Medline: 16648528]

XSL•FO
RenderX

3. Bach B, Simonsen S. How does level of personality functioning inform clinical management and treatment? Implications for ICD-11 classification of personality disorder severity. Curr Opin Psychiatry 2021 Jan;34(1):54-63. [doi: 10.1097/YCO.0000000000000658] [Medline: 33252430]
4. Barnicot K, Redknap C, Coath F, Hommel J, Couldrey L, Crawford M. Patient experiences of therapy for borderline personality disorder: commonalities and differences between dialectical behaviour therapy and mentalization-based therapy and relation to outcomes. Psychol Psychother 2022 Mar;95(1):212-233. [doi: 10.1111/papt.12362] [Medline: 34459086]
5. Deisenhofer AK, Barkham M, Beierl ET, Schwartz B, Aafjes-van Doorn K, Beevers CG, et al. Implementing precision methods in personalizing psychological therapies: barriers and possible ways forward. Behav Res Ther 2024 Jan;172:104443 [FREE Full text] [doi: 10.1016/j.brat.2023.104443] [Medline: 38086157]
6. Delamain H, Buckman JE, O'Driscoll C, Suh JW, Stott J, Singh S, et al. Predicting post-treatment symptom severity for adults receiving psychological therapy in routine care for generalised anxiety disorder: a machine learning approach. Psychiatry Res 2024 Jun;336:115910 Article number 115910 [FREE Full text] [doi: 10.1016/j.psychres.2024.115910] [Medline: 38608539]
7. Stewart R. The big case register. Acta Psychiatr Scand 2014 Aug;130(2):83-86. [doi: 10.1111/acps.12279] [Medline: 24730985]
8. Crawford MJ, Price K, Gordon F, Josson M, Taylor B, Bateman A, et al. Engagement and retention in specialist services for people with personality disorder. Acta Psychiatr Scand 2009 Apr;119(4):304-311. [doi: 10.1111/j.1600-0447.2008.01306.x] [Medline: 19120048]
9. Hansen SJ, Christensen S, Kongerslev MT, First MB, Widiger TA, Simonsen E, et al. Mental health professionals' perceived clinical utility of the ICD-10 vs. ICD-11 classification of personality disorders. Personal Ment Health 2019 May;13(2):84-95. [doi: 10.1002/pmh.1442] [Medline: 30989832]
10. Siefert CJ, Porcerelli J, Meehan KB, Dauphin B. Clinical utility of proposed dimensional personality disorder models: considerations and opportunities. In: Huprich SK, editor. Personality Disorders and Pathology: Integrating Clinical Assessment and Practice in the DSM-5 and ICD-11 Era. New York, NY: American Psychological Association; 2022:53-73.
11. Bach B, Kramer U, Doering S, di Giacomo E, Hutsebaut J, Kaera A, et al. The ICD-11 classification of personality disorders: a European perspective on challenges and opportunities. Borderline Personal Disord Emot Dysregul 2022 Apr 01;9(1):12-63 [FREE Full text] [doi: 10.1186/s40479-022-00182-0] [Medline: 35361271]
12. Campbell K, Lakeman R. Borderline personality disorder: a case for the right treatment, at the right dose, at the right time. Issues Ment Health Nurs 2021 Jun;42(6):608-613. [doi: 10.1080/01612840.2020.1833119] [Medline: 33090931]
13. Lau P, Amestoy ME, Roth M, Monson C. Patient-related factors associated with patient retention and non-completion in psychosocial treatment of borderline personality disorder: a systematic review. Personal Ment Health 2024 Nov;18(4):300-322. [doi: 10.1002/pmh.1627] [Medline: 38807472]
14. Sampson M, Berry C, Jerry A, Gray C. Models of SCM implementation— United Kingdom. In: Mitchell S, Sampson M, Bateman A, editors. Structured Clinical Management (SCM) for Personality Disorder: An Implementation Guide. Oxford, UK: Oxford University Press; 2022:87-112.
15. Lamph G, Baker J, Dickinson T, Lovell K. Personality disorder co-morbidity in primary care 'Improving Access to Psychological Therapy' (IAPT) services: a qualitative study exploring patient perspectives on treatment experience. Behav Cogn Psychother 2021 Mar;49(2):144-158 [FREE Full text] [doi: 10.1017/S1352465820000594] [Medline: 32895075]
16. Fok ML, Stewart R, Hayes RD, Moran P. The impact of co-morbid personality disorder on use of psychiatric services and involuntary hospitalization in people with severe mental illness. Soc Psychiatry Psychiatr Epidemiol 2014 Oct;49(10):1631-1640 [FREE Full text] [doi: 10.1007/s00127-014-0874-4] [Medline: 24715236]
17. Lutz W, Deisenhofer AK, Rubel J, Bennemann B, Giesemann J, Poster K, et al. Prospective evaluation of a clinical decision support system in psychological therapy. J Consult Clin Psychol 2022 Jan;90(1):90-106. [doi: 10.1037/ccp0000642] [Medline: 34166000]
18. Bonfils KA, Dreison KC, Luther L, Fukui S, Dempsey AE, Rapp CA, et al. Implementing CommonGround in a community mental health center: lessons in a computerized decision support system. Psychiatr Rehabil J 2018 Sep;41(3):216-223 [FREE Full text] [doi: 10.1037/prj0000225] [Medline: 27732033]
19. Dawoodbhoy FM, Delaney J, Cecula P, Yu J, Peacock I, Tan J, et al. AI in patient flow: applications of artificial intelligence to improve patient flow in NHS acute mental health inpatient units. Heliyon 2021 May;7(5):e06993 e06993 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06993] [Medline: 34036191]
20. Rose S, McGuire TG. Limitations of P-values and R-squared for stepwise regression building: a fairness demonstration in health policy risk adjustment. Am Stat 2019;73(Suppl 1):152-156 [FREE Full text] [doi: 10.1080/00031305.2018.1518269] [Medline: 31263291]
21. Mathlin G, Freestone M, Taylor C, Shaw J. Offenders with personality disorder who fail to progress: a case-control study using partial least squares structural equation modeling path analysis. JMIRx Med 2021 Oct 29;2(4):e27907 [FREE Full text] [doi: 10.2196/27907] [Medline: 37725548]
22. Ebrahimi A, Poursharifi H, Dolatshahi B, Rezaee O, Hassanabadi HR, Naeem F. The cognitive model of negative symptoms in schizophrenia: a hierarchical component model with PLS-SEM. Front Psychiatry 2021;12:707291 [FREE Full text] [doi: 10.3389/fpsyt.2021.707291] [Medline: 34366940]

XSL•FO
RenderX

23. Ostovar S, Bagheri R, Griffiths MD, Mohd Hashima IH. Internet addiction and maladaptive schemas: the potential role of disconnection/rejection and impaired autonomy/performance. Clin Psychol Psychother 2021 Nov 23;28(6):1509-1524. [doi: 10.1002/cpp.2581] [Medline: 33687117]
24. Saunders R, Buckman JE, Pilling S. Latent variable mixture modelling and individual treatment prediction. Behav Res Ther 2020 Jan;124:103505 [FREE Full text] [doi: 10.1016/j.brat.2019.103505] [Medline: 31841709]
25. Wasserstein RL, Schirm AL, Lazar NA. Moving to a world beyond "p < 0.05". Am Stat 2019 Mar 20;73(sup1):1-19. [doi: 10.1080/00031305.2019.1583913]
26. Amiot O, Sauvaget A, Alamome I, Bulteau S, Charpeaud T, Clair A, et al. Prospective acceptability of digital therapy for major depressive disorder in France: multicentric real-life study. JMIR Form Res 2024 May 20;8:e53204 [FREE Full text] [doi: 10.2196/53204] [Medline: 38568139]
27. Friesen L, Gaine G, Klaver E, Klingle K, Parmar D, Hrabok M, et al. Bridging the gap in community care for patients with borderline personality disorder: protocol for qualitative inquiry into patient, caregiver, and clinician perspectives on service gaps and potential solutions for severe emotion dysregulation. JMIR Res Protoc 2020 Aug 20;9(8):e14885 [FREE Full text] [doi: 10.2196/14885] [Medline: 32815818]
28. Stringer B, van Meijel B, Karman P, Koekkoek B, Hoogendoorn AW, Kerkhof AJ, et al. Collaborative care for patients with severe personality disorders: preliminary results and active ingredients from a pilot study (Part I). Perspect Psychiatr Care 2015 Jul;51(3):180-189. [doi: 10.1111/ppc.12079] [Medline: 25088207]
29. Singham T, Saunders R, Brooker H, Creese B, Aarsland D, Hampshire A, et al. Are subtypes of affective symptoms differentially associated with change in cognition over time: a latent class analysis. J Affect Disord 2022 Jul 15;309:437-445 [FREE Full text] [doi: 10.1016/j.jad.2022.04.139] [Medline: 35490883]
30. Cole VT, Bauer DJ. A note on the use of mixture models for individual prediction. Struct Equ Modeling 2016;23(4):615-631 [FREE Full text] [doi: 10.1080/10705511.2016.1168266] [Medline: 27346932]
31. Tueller S, Lubke G. Evaluation of structural equation mixture models parameter estimates and correct class assignment. Struct Equ Modeling 2010 Apr 01;17(2):165-192 [FREE Full text] [doi: 10.1080/10705511003659318] [Medline: 20582328]
32. Berlin KS, Williams NA, Parra GR. An introduction to latent variable mixture modeling (part 1): overview and cross-sectional latent class and latent profile analyses. J Pediatr Psychol 2014 Mar;39(2):174-187. [doi: 10.1093/jpepsy/jst084] [Medline: 24277769]
33. Vermunt JK, Magidson J. Latent class cluster analysis. In: Hagenaars JA, McCutcheon AL, editors. Applied Latent Class Analysis. Cambridge, MA: Cambridge University Press; 2002:89-106.
34. McLachlan GJ, Lee SX, Rathnayake SI. Finite mixture models. Annu Rev Stat Appl 2019 Mar 07;6(1):355-378. [doi: 10.1146/annurev-statistics-031017-100325]
35. Scrucca L, Fop M, Murphy TB, Raftery AE. mclust 5: clustering, classification and density estimation using gaussian finite mixture models. R J 2016 Aug;8(1):289-317 [FREE Full text] [Medline: 27818791]
36. Newton-Howes G, Cunningham R, Atkinson J. Personality disorder prevalence and correlates in a whole of nation dataset. Soc Psychiatry Psychiatr Epidemiol 2021 Apr;56(4):679-685. [doi: 10.1007/s00127-020-01876-y] [Medline: 32394007]
37. Bell D. Delivering personalised cancer care to enhance patients' quality of life. Cancer Nurs Pract 2022;21(5):52. [doi: 10.7748/cnp.2022.e1790]
38. Fonagy P, Allison E. The role of mentalizing and epistemic trust in the therapeutic relationship. Psychotherapy (Chic) 2014 Sep;51(3):372-380. [doi: 10.1037/a0036505] [Medline: 24773092]
39. Nylund-Gibson K, Grimm RP, Masyn KE. Prediction from latent classes: a demonstration of different approaches to include distal outcomes in mixture models. Struct Equ Modeling 2019 Apr 24;26(6):967-985. [doi: 10.1080/10705511.2019.1590146]
40. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
41. Bensmail H, Celeux G. Regularized gaussian discriminant analysis through eigenvalue decomposition. J Am Stat Assoc 1996 Dec;91(436):1743-1748. [doi: 10.1080/01621459.1996.10476746]
42. Masyn KE. Latent class analysis and finite mixture modeling. In: Little TD, editor. The Oxford Handbook of Quantitative Methods in Psychology: Vol. 2: Statistical Analysis. Oxford, UK: Oxford Academic Press; 2013:551-611.
43. Saunders R, Cape J, Fearon P, Pilling S. Predicting treatment outcome in psychological treatment services by identifying latent profiles of patients. J Affect Disord 2016 Jun;197:107-115 [FREE Full text] [doi: 10.1016/j.jad.2016.03.011] [Medline: 26991365]
44. Colling C, Khondoker M, Patel R, Fok M, Harland R, Broadbent M, et al. Predicting high-cost care in a mental health setting. BJPsych Open 2020 Jan 17;6(1):e10 [FREE Full text] [doi: 10.1192/bjo.2019.96] [Medline: 31950891]
45. Shmueli G, Ray S, Velasquez Estrada JM, Chatla SB. The elephant in the room: predictive performance of PLS models. J Bus Res 2016 Oct;69(10):4552-4564. [doi: 10.1016/j.jbusres.2016.03.049]
46. Takano K, Hino H, Akaho S, Murata N. Nonparametric e-mixture estimation. Neural Comput 2016 Dec;28(12):2687-2725. [doi: 10.1162/NECO_a_00888] [Medline: 27626969]
47. Grimm KJ, Houpt R, Rodgers D. Model fit and comparison in finite mixture models: a review and a novel approach. Front Educ 2021 Mar 4;6:25. [doi: 10.3389/feduc.2021.613645]

XSL•FO
RenderX

48. Melnykov V. On the distribution of posterior probabilities in finite mixture models with application in clustering. J Multivar Anal 2013 Nov;122:175-189. [doi: 10.1016/j.jmva.2013.07.014]
49. Perera G, Broadbent M, Callard F, Chang C, Downs J, Dutta R, et al. Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) case register: current status and recent enhancement of an electronic mental health record-derived data resource. BMJ Open 2016 Mar 01;6(3):e008721 [FREE Full text] [doi: 10.1136/bmjopen-2015-008721] [Medline: 26932138]
50. Fernandes AC, Cloete D, Broadbent MT, Hayes RD, Chang C, Jackson RG, et al. Development and evaluation of a de-identification procedure for a case register sourced from mental health electronic records. BMC Med Inform Decis Mak 2013 Jul 11;13:71 [FREE Full text] [doi: 10.1186/1472-6947-13-71] [Medline: 23842533]
51. Ranga R. South London and Maudsley discharge & transfer policy. National Health Service, England. URL: https://slam.nhs.uk/download.cfm?doc=docm93jijm4n1241.pdf&ver=1721 [accessed 2021-11-26]
52. Tyrer P. Response: Personality disorder is here to stay across the lifespan - a response to Commentaries on the May 2022 Debate. Child Adolesc Ment Health 2022 Sep;27(3):253-255. [doi: 10.1111/camh.12583] [Medline: 35842921]
53. Pan B, Wang W. Practical implications of ICD-11 personality disorder classifications. BMC Psychiatry 2024 Mar 07;24(1):191-202 [FREE Full text] [doi: 10.1186/s12888-024-05640-3] [Medline: 38454364]
54. Coulacoglou C, Saklofske DH. Advances in latent variable measurement modeling. In: Coulacoglou C, Saklofske DH, editors. Psychometrics and Psychological Assessment: Principles and Applications. New York, NY: Academic Press; 2017:67-88.
55. Katakis P, Schlief M, Barnett P, Rains LS, Rowe S, Pilling S, et al. Effectiveness of outpatient and community treatments for people with a diagnosis of 'personality disorder': systematic review and meta-analysis. BMC Psychiatry 2023 Jan 21;23(1):57-86 [FREE Full text] [doi: 10.1186/s12888-022-04483-0] [Medline: 36681805]
56. Steadman J. Mapping the complex care pathways for personality disorder in South London (in press). Dissertation. Queen Mary, University of London. 2025 Aug 12. URL: https://qmro.qmul.ac.uk/xmlui/handle/123456789/98877 [accessed 2025-03-12]
57. McLennan D, Noble S, Plunkett E, Wright G, Gutacker N. The English indices of deprivation. Ministry of Housing CLG. 2019. URL: https://www.gov.uk/government/collections/english-indices-of-deprivation [accessed 2024-04-29]
58. Cunningham H. GATE: a framework and graphical development environment for robust NLP tools and applications. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics. 2002 Presented at: ACL '02; July 6-12, 2002; Philadelphia, PA p. 33-86. [doi: 10.3115/1073083.1073112]
59. The ICD-10 classification of mental and behavioural disorders. World Health Organization. 1993. URL: https://icd.who.int/browse10/2019/en [accessed 2024-04-29]
60. Tyrer P, Mulder R, Kim YR, Crawford MJ. The development of the ICD-11 classification of personality disorders: an amalgam of science, pragmatism, and politics. Annu Rev Clin Psychol 2019 May 07;15:481-502. [doi: 10.1146/annurev-clinpsy-050718-095736] [Medline: 30601688]
61. Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. J Abnorm Psychol 2017 May;126(4):454-477. [doi: 10.1037/abn0000258] [Medline: 28333488]
62. Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Struct Equ Modeling 2007 Dec 05;14(4):535-569. [doi: 10.1080/10705510701575396]
63. Almodóvar-Rivera IA, Pericchi-Guerra LR. An objective and robust bayes factor for the hypothesis test one sample and two population means. Entropy (Basel) 2024 Jan 20;26(1):e26010088 [FREE Full text] [doi: 10.3390/e26010088] [Medline: 38275496]
64. Rosenberg JM, Beymer PN, Anderson DJ, van Lissa CJ, Schmidt JA. tidyLPA: an R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. J Open Source Softw 2018 Oct;3(30):978. [doi: 10.21105/joss.00978]
65. Bray BC, Lanza ST, Tan X. Eliminating bias in classify-analyze approaches for latent class analysis. Struct Equ Modeling 2015 Jan;22(1):1-11 [FREE Full text] [doi: 10.1080/10705511.2014.935265] [Medline: 25614730]
66. Feingold A, Tiberio SS, Capaldi DM. New approaches for examining associations with latent categorical variables: applications to substance abuse and aggression. Psychol Addict Behav 2014 Mar;28(1):257-267 [FREE Full text] [doi: 10.1037/a0031487] [Medline: 23772759]
67. Celeux G, Soromenho G. An entropy criterion for assessing the number of clusters in a mixture model. J Classif 1996 Sep;13(2):195-212. [doi: 10.1007/BF01246098] [Medline: 30311153]
68. McLachlan GJ. Discriminant analysis. Wiley Interdiscip Rev Comput Stat 2012 Jul 16;4(5):421-431. [doi: 10.1002/wics.1219]
69. Baudry JP, Raftery AE, Celeux G, Lo K, Gottardo R. Combining mixture components for clustering. J Comput Graph Stat 2010 Jun 01;9(2):332-353 [FREE Full text] [doi: 10.1198/jcgs.2010.08111] [Medline: 20953302]
70. Hair Jr JF, Hult GT, Ringle CM, Sarstedt M, Danks NP, Ray S. Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R: A Workbook. Cham, Switzerland: Springer; 2021.
71. Hamner B, Frasco M, LeDell E. Metrics. GitHub. URL: https://github.com/mfrasco/Metrics [accessed 2024-04-29]

XSL•FO
RenderX

72. Khair U, Fahmi H, Al Hakim S, Rahim R. Forecasting error calculation with mean absolute deviation and mean absolute percentage error. J Phys Conf Ser 2017 Dec 14;930:012002. [doi: 10.1088/1742-6596/930/1/012002]
73. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
74. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/ [accessed 2024-04-29]
75. Integrated development for R. RStudio. URL: http://www.rstudio.com/ [accessed 2024-04-29]
76. DeLeo K, Maconick L, McCabe R, Broeckelmann E, Sheridan Rains L, Rowe S, et al. Experiences of crisis care among service users with complex emotional needs or a diagnosis of 'personality disorder', and other stakeholders: systematic review and meta-synthesis of the qualitative literature. BJPsych Open 2022 Feb 24;8(2):e53 [FREE Full text] [doi: 10.1192/bjo.2022.1] [Medline: 35197131]
77. Lo Y, Mendell NR, Rubin DB. Testing the number of components in a normal mixture. Biometrika 2001 Oct 01;88(3):767-778. [doi: 10.1093/biomet/88.3.767]
78. McLachlan GJ. On bootstrapping the likelihood ratio test stastistic for the number of components in a normal mixture. R Stat Soc C Appl Stat 1987;36(3):318-324. [doi: 10.2307/2347790]
79. Bolck A, Croon M, Hagenaars J. Estimating latent structure models with categorical variables: one-step versus three-step estimators. Polit Anal 2017 Jan 04;12(1):3-27. [doi: 10.1093/pan/mph001]
80. Nielsen JD, Rosenthal JS, Sun Y, Day DM, Bevc I, Duchesne T. Group-based criminal trajectory analysis using cross-validation criteria. Commun Stat Theory Methods 2014 Sep 30;43(20):4337-4356. [doi: 10.1080/03610926.2012.719986]
81. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence. 1995 Presented at: IJCAI '95; August 20-25, 1995; Montreal, QC p. 1137-1143 URL: https://dl.acm.org/doi/10.5555/1643031.1643047 [doi: 10.5555/1643031.1643047]
82. Brier GW. Verification of forecasts expressed in terms of probability. Mon Weather Rev 1950 Jan;78(1):1-3. [doi: 10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2]
83. Merkle EC, Flores RD. Forecaster evaluation as model comparison. Decision 2024 May 16:43-62. [doi: 10.1037/dec0000230]
84. McMullen L, Parghi N, Rogers ML, Yao H, Bloch-Elkouby S, Galynker I. The role of suicide ideation in assessing near-term suicide risk: a machine learning approach. Psychiatry Res 2021 Oct;304:114118. [doi: 10.1016/j.psychres.2021.114118] [Medline: 34403873]
85. Bennemann B, Schwartz B, Giesemann J, Lutz W. Predicting patients who will drop out of out-patient psychotherapy using machine learning algorithms. Br J Psychiatry 2022 Feb 18;220(4):1-10. [doi: 10.1192/bjp.2022.17] [Medline: 35177132]
86. Shmueli G, Sarstedt M, Hair JF, Cheah J, Ting H, Vaithilingam S, et al. Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. Eur J Mark 2019 Nov 11;53(11):2322-2347. [doi: 10.1108/ejm-02-2019-0189]
87. Hair JF, Astrachan CB, Moisescu OI, Radomir L, Sarstedt M, Vaithilingam S, et al. Executing and interpreting applications of PLS-SEM: updates for family business researchers. J Fam Bus Strateg 2021 Sep;12(3):100392. [doi: 10.1016/j.jfbs.2020.100392]
88. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
89. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007 Oct 20;335(7624):806-808 [FREE Full text] [doi: 10.1136/bmj.39335.541782.AD] [Medline: 17947786]
90. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, RECORD Working Committee. The REporting of studies conducted using observational routinely-collected health data (RECORD) statement: methods for arriving at consensus and developing reporting guidelines. PLoS Med 2015 Oct;12(10):e1001885 [FREE Full text] [doi: 10.1371/journal.pmed.1001885] [Medline: 26440803]
91. van der Nest G, Lima Passos V, Candel MJ, van Breukelen GJ. An overview of mixture modelling for latent evolutions in longitudinal data: modelling approaches, fit statistics and software. Adv Life Course Res 2020 Mar;43:100323. [doi: 10.1016/j.alcr.2019.100323] [Medline: 36726256]
92. Nylund-Gibson K, Masyn KE. Covariates and mixture modeling: results of a simulation study exploring the impact of misspecified effects on class enumeration. Struct Equ Modeling 2016 Sep 22;23(6):782-797. [doi: 10.1080/10705511.2016.1221313]
93. Simola U, Cisewski-Kehe J, Wolpert RL. Approximate Bayesian computation for finite mixture models. J Stat Comput Simul 2020 Nov 06;91(6):1155-1174. [doi: 10.1080/00949655.2020.1843169]
94. Bourke J, Murphy A, Flynn D, Kells M, Joyce M, Hurley J. Borderline personality disorder: resource utilisation costs in Ireland. Ir J Psychol Med 2021 Sep;38(3):169-176. [doi: 10.1017/ipm.2018.30] [Medline: 34465404]
95. Sveen CA, Pedersen G, Ulvestad DA, Zahl KE, Wilberg T, Kvarstein EH. Societal costs of personality disorders among treatment-seeking patients in Norway: the relative contribution of specific DSM-5 categories. Eur Arch Psychiatry Clin Neurosci 2024 Feb;274(1):139-149 [FREE Full text] [doi: 10.1007/s00406-023-01655-1] [Medline: 37598131]

XSL•FO
RenderX

96. Durvasula SR, Halkitis PN. Delineating the interplay of personality disorders and health. Behav Med 2017;43(3):151-155 [FREE Full text] [doi: 10.1080/08964289.2017.1337400] [Medline: 28767019]
97. Bender DS, Skodol AE, Dyck IR, Markowitz JC, Shea MT, Yen S, et al. Ethnicity and mental health treatment utilization by patients with personality disorders. J Consult Clin Psychol 2007 Dec;75(6):992-999. [doi: 10.1037/0022-006X.75.6.992] [Medline: 18085915]
98. Cailhol L, Pelletier É, Rochette L, Renaud S, Koch M, David P, et al. Utilization of health care services by patients with cluster B personality disorders or schizophrenia. Psychiatr Serv 2021 Dec 01;72(12):1392-1399. [doi: 10.1176/appi.ps.202000554] [Medline: 34281361]
99. Finkelstein A, Zhou A, Taubman S, Doyle J. Health care hotspotting - a randomized, controlled trial. N Engl J Med 2020 Jan 09;382(2):152-162 [FREE Full text] [doi: 10.1056/NEJMsa1906848] [Medline: 31914242]
100. Capp R, Misky GJ, Lindrooth RC, Honigman B, Logan H, Hardy R, et al. Coordination program reduced acute care use and increased primary care visits among frequent emergency care users. Health Aff (Millwood) 2017 Oct 01;36(10):1705-1711. [doi: 10.1377/hlthaff.2017.0612] [Medline: 28971914]
101. Botham J, Simpson A, McCrone P. Mental health service use and costs associated with complex emotional needs and a diagnosis of personality disorder: analysis of routine data. BJPsych Bull 2024 Apr;48(2):85-92 [FREE Full text] [doi: 10.1192/bjb.2023.41] [Medline: 37310185]
102. Catalao R, Broadbent M, Ashworth M, Das-Munshi J, L Hatch S, Hotopf M, et al. Access to psychological therapies amongst patients with a mental health diagnosis in primary care: a data linkage study. Soc Psychiatry Psychiatr Epidemiol (Forthcoming) 2024 Nov 06. [doi: 10.1007/s00127-024-02787-y] [Medline: 39503853]
103. Parsons J, Bryce C, Atherton H. Which patients miss appointments with general practice and the reasons why: a systematic review. Br J Gen Pract 2021 Jun;71(707):e406-e412 [FREE Full text] [doi: 10.3399/BJGP.2020.1017] [Medline: 33606660]
104. Talukder AK, Sanz JB, Samajpati J. 'Precision health': balancing reactive care and proactive care through the evidence based knowledge graph constructed from real-world electronic health records, disease trajectories, diseasome, and patholome. In: Proceedings of the 8th International Conference on Big Data Analytics. 2020 Presented at: BDA '20; December 15-18, 2020; Sonepat, India p. 113-133 URL: https://link.springer.com/chapter/10.1007/978-3-030-66665-1_9 [doi: 10.1007/978-3-030-66665-1_9]
105. Population health management in primary health care: a proactive approach to improve health and well-being: primary heath care policy paper series. World Health Organization. 2023. URL: https://apps.who.int/iris/bitstream/handle/10665/368805/WHO-EURO-2023-7497-47264-69316-eng.pdf [accessed 2024-04-29]

Abbreviations

BIC: Bayesian information criterion
CONSORT: Consolidated Standards of Reporting Trials
CRIS: Clinical Records Interactive Search
ICD-10: International Classification of Diseases, Tenth Revision
ICD-11: International Classification of Diseases, Eleventh Revision
LCA: latent class analysis
LP: latent profile
LPA: latent profile analysis
LRT: likelihood ratio test
MAE: mean absolute error
MAPE: mean absolute percentage error
NHS: National Health Service
NIHR: National Institute for Health Research
PD: personality disorder
SLaM: South London and Maudsley

Edited by T de Azevedo Cardoso; submitted 11.12.23; peer-reviewed by P Moran, J Ballard, P Roma; comments to author 01.03.24; revised version received 06.07.24; accepted 10.12.24; published 15.04.25.

Please cite as:
Steadman J, Saunders R, Freestone M, Stewart R
Subtyping Service Receipt in Personality Disorder Services in South London: Observational Validation Study Using Latent Profile Analysis
Interact J Med Res 2025;14:e55348
URL: https://www.i-jmr.org/2025/1/e55348
doi:10.2196/55348
PMID:

XSL•FO
RenderX

©Jack Steadman, Rob Saunders, Mark Freestone, Robert Stewart. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 15.04.2025. 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

Detection of Chronic Cognitive-Motor Deficits in Adults With a History of Concussion Using Computerized Eye-Hand Coordination Tasks: Preliminary Experimental Design Study

Qin Zhu[1*], PhD; Shaochen Huang[2*], PhD

[1]Division of Kinesiology and Health, University of Wyoming, Laramie, WY, United States

[2]School of Health Promotion and Kinesiology, Texas Woman's University, 1600 N Bell Ave, Pioneer Hall, Denton, TX, United States

[*]all authors contributed equally

Corresponding Author:
Shaochen Huang, PhD
School of Health Promotion and Kinesiology, Texas Woman's University, 1600 N Bell Ave, Pioneer Hall, Denton, TX, United States

Abstract

Background: Concussion has been a major public health concern due to the substantial cognitive sequelae it results. Although the dysfunctions of the frontal lobe and corpus callosum owing to concussions have been documented, the existing concussion screening tools merely examine cognitive functions in isolation of motor functions and failed to detect the chronic cognitive-motor impairments following concussions. Yet, there has been no concussion screening test aimed specifically to assess the coupled cognitive-motor functions.

Objective: This study aimed to provide preliminary evidence for using computerized eye-hand coordination tasks to detect chronic cognitive-motor deficits associated with concussion history.

Methods: The computerized eye-hand coordination tasks were used to assess the coupled cognitive-motor functions of the participants with and with no history of concussion. In experiment 1, a total of 12 participants (6 young adults with a history of concussion and 6 healthy controls) completed longitudinal assessments of coordination profiles across a year. Experiment 2 examined a total of 20 participants (10 participants with a history of concussion and 10 healthy controls) using an iterated single-session protocol. *Just noticeable difference* (JND) and *proportion of time-on-task* (PTT) were used to assess cognitive-motor performance. Mixed-design ANOVAs were used to examine group differences, and the effect sizes were assessed using Cohen *d* test.

Results: In experiment 1, participants with a history of concussion exhibited more inconsistent ability to visually discriminate the in-phase coordination pattern (coefficient of variation of JND: participants with a history of concussion = mean 0.27, SD 0.04, and healthy controls = mean 0.17, SD 0.07; t_{10}=2.93; *P*=.02). Similarly, their performance on unimanual and bimanual in-phase and anti-phase coordination patterns was significantly poorer (at in-phase: $PTT_{Concussed}$=mean 0.63, SD 0.10, and $PTT_{Healthy}$=mean 0.73, SD 0.08 [$F_{1,10}$=8.49; *P*=.02]; at anti-phase: $PTT_{Concussed}$=mean 0.46, SD 0.14, and $PTT_{Healthy}$=mean 0.60, SD 0.10 [$F_{1,10}$=10.67; *P*=.008]). In experiment, 2 where only the unimanual coordination tasks were implemented for screening, participants with a history of concussion showed impaired performance in both in-phase and anti-phase tasks (at in-phase: $PTT_{Concussed}$=mean 0.62, SD 0.13, and $PTT_{Healthy}$=mean 0.74, SD 0.07 [$F_{1,54}$=4.20; *P*=.045]; at anti-phase: $PTT_{Concussed}$=mean 0.37, SD 0.15, and $PTT_{Healthy}$=mean 0.56, SD 0.14 [$F_{1,54}$=10.26; *P*=.002]), and they also failed to show the differentiated performance between anti-phase and 90° coordination patterns ($PTT_{Anti\text{-}phase}$=mean 0.37, SD 0.15, and $PTT_{90°\ coordination}$=mean 0.37, SD 0.13).

Conclusions: Due to their ability to detect both impaired and undifferentiated performance in producing intrinsic and novel coordination patterns, the unimanual coordination tasks appear to be a sensitive screening tool for chronic cognitive-motor deficits associated with history of concussion.

(Interact J Med Res 2025;14:e70867) doi:10.2196/70867

KEYWORDS

concussion screening; coordination; cognitive-motor functions; perception-action coupling; eye-hand coordination

XSL•FO
RenderX

Introduction

Most concussions fall in the category of mild traumatic brain injury (mTBI) [1,2]. More than 81% of the 69 million traumatic brain injuries that occur annually worldwide are considered mTBI [3]. Despite the label of "mild," individuals with concussion often experience cognitive, physical, psychological, and social dysfunctions that severely impact the quality of their lives, causing huge social and economic burdens [4-6]. Although individuals who experienced concussions could recover spontaneously within a few weeks to months, persisting symptoms and permanent deficits could remain as severe sequelae [7]. Thus, the National Institute of Neurological Disorders and Stroke and the National Institutes of Health declared concussion as a major public health concern, and efforts to improve early diagnosis and reduce postconcussion disability should be one of the national research priorities [8,9].

The frontal lobe is known as a brain area where the primary motor cortex (M1) is located and responsible for both cognitive and motor functions (eg, attention, memory, action planning, and decision-making). It has been reported that the functions of the frontal lobe are significantly impaired following a severe or moderate traumatic brain injury [10,11] or with enduring multiple mTBI [12,13]. Although brain atrophy could result from multiple concussions [14], clinical neuroimaging is not routinely performed for a reported case of concussion due to the high cost and the likelihood of undetected brain structural damage [15,16]. Therefore, numerous neurobehavioral assessment tools have been developed to examine the frontal lobe dysfunctions associated with concussion. For instance, the Sport Concussion Assessment Tool is a standardized screening tool that has been widely adopted for immediate assessment of a sport-related concussion, which consists of a series of tasks to assess cognitive (eg, attention and memory) and motor (eg, movement speed or accuracy and balance) functions of athletes. Although effective, the existing screening tools of concussion have two limitations. First, they are insensitive to capture the subacute and chronic cognitive-motor impairments following concussions. A scoping review [17] showed that a large proportion of individuals with a single concussion continued to demonstrate the measurable cognitive impairment, whereas the standard neurobehavioral assessments indicated no cognitive impairments. Accordingly, Sport Concussion Assessment Tool 5th edition has been recommended to be inappropriate for assessing people in the subacute and chronic phases of injury [18]. Second, they examine the cognitive functions in isolation with the motor functions, which consequently increases testing time and decreases efficiency due to the need of integrating different testing scores for overall rating. Given the high integration of cognitive and motor functions in the frontal lobe [19], assessing the coupled cognitive-motor functions should be more straightforward and efficient than the uncoupled approaches, which unfortunately have been missing in the current neurobehavioral screen tools for concussion.

Coordination is required for many functional motor tasks such as walking and running. The brain areas involved in performing and learning coordination tasks are the frontal lobe and cerebellum [20,21]. Research suggests an important link between coordination capability and cognitive-motor functions. Young adults diagnosed with developmental coordination disorder often demonstrate deficits in performing tasks that demand attention, working memory, or cognitive flexibility [22]. Due to its high relevance to cognitive-motor functions, coordination tasks including finger-to-nose test [23], the visuomotor tracking task [24], the divided-attention gait task [25], and bimanual coordination task [26] are often included in neurobehavioral screen tools for concussion. Meanwhile, many studies [27,28] have reported that the brain's largest white matter structure, corpus callosum (CC), is vulnerable to concussion. Since CC is known for its function in interhemispheric processing, the structural damage of CC from concussions is associated with the impaired coordination performance demonstrated by the participants with a history of concussion, which makes coordination task alone as a decent screening tool for concussion. However, in using coordination tasks to assess cognitive-motor dysfunctions associated with concussion, the coordination task is often treated as a motor task with a variety of performance outcome measures (eg, speed, accuracy, motion asymmetry, motion continuity, or pattern stability). In view of the perception-action theory of coordination [29,30], the ability to coordinate limb movements is coupled with the ability to perceive the coordination pattern. Therefore, a comprehensive coordination assessment should involve both performing and perceiving the coordination pattern. Yet, there has been no concussion screening test aimed to assess the coupled cognitive-motor functions associated with coordination.

In this study, we explored whether the computerized eye-hand coordination tasks, implemented to assess the coupled cognitive-motor functions, could be used to detect the chronic and subacute cognitive-motor deficits associated with history of concussion. Three different coordination tasks were implemented through human-computer interaction to assess people's ability to both perform and perceive different coordination patterns. Coordination performance was assessed by the established measures in psychophysical studies [31-33]. Adult participants with a history of concussion were recruited and gender-matched with healthy controls in 2 experiments, and their coordination performance was compared to determine whether a specific coordination task has the power to detect the group difference. Three coordination tasks (visual discrimination, unimanual, and bimanual) were implemented in experiment 1 to detect the group difference longitudinally and determine the sensibility of task and task measures for screening. We hypothesized that the impaired cognitive-motor functions could be detected by the respective task measures.

Experiment 2 was then followed to determine whether a simplified coordination screening test was sufficient to differentiate people with a history of concussion from those with no history of concussion in a single testing session. Based on the findings from experiment 1, we adjusted the testing protocol to assess participants' performance of 3 coordination patterns (in-phase, 90°, and anti-phase) in just 1 testing session. Each pattern was examined by a bout of five 20-second-long trials. Such an adjustment was made to examine whether the unimanual coordination task could serve as a quick and efficient screening tool for concussion. It was hypothesized that the

XSL•FO
RenderX

performance of 3 coordination patterns would differ between participants with a history of concussion and healthy participants.

Methods

Experiment 1

Participants

Six young adults with a history of concussion (mean age 22.00, SD 4.34 years; mean time from last concussion 38.67, SD 31.94 months: 2 males and 4 females) and 6 gender-matched healthy controls with no history of concussion (mean age 20.83, SD 2.64 years) were recruited on campus at the University of Wyoming. All concussions were medically diagnosed with accessible records. All participants had no history of substance abuse (drugs, alcohol, or controlled medication), peripheral neuropathy (including loss or decreased sensation or motor function in the upper extremity), or any neurological disorders that could impair coordination abilities (such as ataxia, giant axonal neuropathy, or multiple sclerosis).

Ethical Considerations

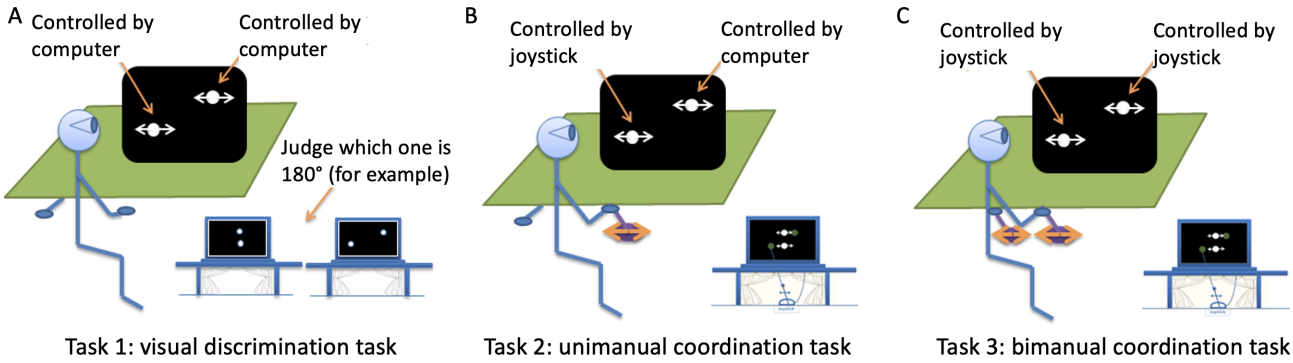
This study was approved by the institutional review board at the University of Wyoming (protocol number 20170220SH01461). All participants provided written informed consent prior to participation and were informed of their right to withdraw from the study at any time without penalty. All data collected were deidentified prior to analysis to ensure participant privacy and confidentiality. No personally identifiable information was stored. No monetary compensation was provided to participants for their participation in this study.

Apparatus

A joystick-computer interaction system was used for implementation of the visual discrimination, unimanual, and bimanual coordination tasks. Participants sat on a stool that could be adjusted for height to see a laptop screen at their eye-height level. A 15" PC laptop with a screen resolution of 1024×768 and a refresh rate of 60 Hz was set on top of a custom-built shelf facing the participant. Two Logitech Force 3D joysticks were connected via USB to the PC laptop, one opposite the left shoulder of the participant and the other opposite the right shoulder. Participants manipulated both joysticks in performing the bimanual task, while only the right joystick was used in performing the unimanual tasks. None of the joysticks was used for the visual discrimination task. The joysticks were hidden underneath by a tablecloth covering the shelf. All participants were able to reach and grasp the joysticks comfortably without seeing them.

The computer displayed 2 white dots against a black background on screen, one above the other. In the visual discrimination task, both dots were controlled by the computer, and participants were asked to watch the computer displays and make the judgments by pressing the assigned keys. In the unimanual task, participants controlled only the bottom dot via the right joystick, while the top dot was controlled by the computer. Finally, in the bimanual coordination task, the top dot was controlled by the left joystick and the bottom by the right joystick all the time (see Figures 1A-1C for experimental setting and tasks). The movement amplitude of the dots was 300 pixels, and they were each 60 pixels in diameter. The viewing distance was 70 cm, yielding an oscillating movement spanned at approximately 7.5° visual angle, equivalent to the movement amplitude of 18 cm in the horizontal plane on the joysticks. Stimulus presentation, data recording, and all data analyses were handled by a custom MATLAB toolbox written by ADW incorporating the Psychtoolbox [34].

Figure 1. An illustration of experimental setting for (A) visual discrimination task; (B) unimanual coordination tasks, and (C) bimanual coordination tasks. Visual feedback about the correctness of the produced pattern was provided by changing the color of dots to green in unimanual and bimanual tasks.



Procedure

All participants were assessed by the following 3 eye-hand coordination tasks in a year period with 6 assessments separated by 2 months. The experimental procedures were designed to assess participants' visual perception and motor coordination abilities.

Visual Discrimination Task

Participants performed a 2-alternative forced choice judgment task with a computer. Each trial started with a 4-second visual demonstration of a target rhythmic pattern (2 dots kept oscillating to one another) on the screen, followed by a pair of 4-second visual displays of rhythmic patterns for judgment: one being the target pattern and the other the distracting pattern. Both target and judgment patterns were played at 0.75 Hz.

XSL•FO
RenderX

Participants must tell which 1 of the 2 presented judgment patterns was the target pattern by pressing a key: "A" for the first presented pattern, and "L" for the second presented pattern. Two intrinsic coordination patterns (in-phase and anti-phase) were used as target patterns. The in-phase pattern showed 2 dots moving in the same direction at the same speed, while the anti-phase pattern showed 2 dots moving in the opposite direction at the same speed, both in the top-bottom configuration. The number of trials was determined using 2 independent but interleaved transformed staircase procedures (see details about the staircase procedure in section "Data Analysis"). Thus, the total number of trials and duration of the task varied for each participant.

Unimanual Coordination Task

Participants performed 36 trials of unimanual joystick task with the joystick-computer interaction system, including 3 trials at the in-phase pattern, 3 trials at the anti-phase pattern, and 30 trials at a novel coordination pattern (90° relative phase, ie, half of the time the 2 dots move in the same direction and half of the time they move in the opposite direction). For each pattern, participants started with a 5-second visual demonstration of a target coordination pattern at 0.75 Hz, followed by trials (10 seconds each) in which they controlled 1 joystick to move 1 dot on the screen together with another dot controlled by the computer to replicate the target coordination pattern. Visual feedback about the correctness of the produced 90° coordination pattern was provided by changing the color of dots to green.

Bimanual Coordination Task

Similar procedures were followed as in the unimanual coordination task. However, instead of controlling 1 joystick, participants controlled 2 joysticks to move the 2 dots on the screen in order to replicate the target coordination pattern. Two intrinsic coordination patterns (ie, in-phase and anti-phase) and the novel coordination pattern were assessed with visual feedback provided only for the latter.

In this study, we were specifically interested in participants' ability to learn a novel coordination pattern with visual feedback. As numerous studies have shown that in-phase and anti-phase are intrinsic coordination patterns that healthy people can produce readily [35,36], while 90° coordination is challenging and requires a significant amount of training to produce [37], we provided more trials and visual feedback about the correctness for participants to perform 90° coordination in both unimanual and bimanual tasks.

Data Analysis

The staircase method was used to calculate the *just noticeable difference* (JND) in the visual discrimination task. The staircase is a psychophysical procedure for estimating perceptual thresholds, which makes the judgments harder or easier as a function of the correctness of the last trials. In this study, the magnitude of distracting patterns (the difference between the 2 displayed coordination patterns) was determined by a 1-up or 2-down rule defined in the staircase method. Considering the time constraint for a screening task and based on a previous study on the perception of coordination patterns [38], we applied a step size "up" of 1° and a stop rule of 3 reversals (errors) to the in-phase, and a step size "up" of 5° and a stop rule of 2 reversals (errors) to the anti-phase. Step size "down" was fixed based on Table 5.1 of Kingdom and Prins [39]. Thus, it was 0.55° for in-phase and 2.74° for anti-phase. The initial difference was set to 5° and 40° for in-phase and anti-phase, respectively. Each time the trial stepped down until the first error occurred, then it repeated until meeting the stop rule. In addition, the *coefficient of variation* (CV) was calculated for JNDs collected from the 6 assessments, respectively, for in-phase and anti-phase patterns to examine how consistent the participant was able to discriminate intrinsic coordination patterns.

In unimanual and bimanual coordination tasks, the *proportion of time-on-task* (PTT) was automatically calculated by the system and used to quantify the coordination performance. Specifically, the proportion of each continuous relative phase time series (trial) that fell within the range of the target phase ±20° tolerance was computed. PTT has been proven as a valid measure of performance at the required relative phase that can capture both accuracy and stability of coordination performance in numerous studies [32,33,40].

To examine the group difference in each of the eye-hand coordination tasks across the 6 testing sessions, the mixed-design ANOVA was performed on JNDs in the visual discrimination task and mean PTTs in both unimanual and bimanual coordination tasks, respectively, treating group (group with a history of concussion vs healthy controls) as a between-subject variable and testing sessions (1 through 6) as a within-subject variable. Prior to conducting ANOVAs, the normality of outcome variables was assessed using the Shapiro-Wilk test and all variables met the normality assumption. For post hoc pairwise comparisons following the detection of significant ANOVA effect, we applied Tukey HSD (Honestly Significant Difference) corrections to adjust for multiple comparisons. Since the study purpose was to examine the use of eye-hand coordination tasks to detect the postconcussion cognitive-motor deficits, we were more interested in the main effect of group, and the main effect of testing sessions would indicate only the learning effect associated with participants being familiar with the assessment tasks. In addition, the 2-tailed independent samples *t* test was performed on CV of JND at in-phase and anti-phase patterns, respectively.

Finally, due to the small sample size that might have limited the power of the study to detect the effects of interest, Cohen *d* effect size was calculated for each investigated factor. According to Cohen [41], d=0.20, d=0.50, and d=0.80 were considered as "small," "medium," and "large" effect sizes, respectively. A large effect size would indicate a strong effect to be detected in the future by a larger sample size.

Experiment 2

Participants

Another 10 young adults with a history of concussion (mean age 20.90, SD 1.45 years; mean time from last concussion 23.88, SD 17.04 months: 6 males and 4 females) and 10 gender-matched healthy controls (mean age 22.30, SD 1.70 years) were recruited on campus at the University of Wyoming.

XSL•FO
RenderX

The inclusion and exclusion criteria were identical to those used in the first experiment.

Ethical Considerations

This study was approved by the institutional review board at the University of Wyoming (protocol number 20180425SH01973). The same informed consent process, privacy protections, and compensation policies applied as in experiment 1.

Apparatus and Procedure

Only the unimanual coordination task was used, and participants were tested with 3 coordination patterns (in-phase, anti-phase, and 90° coordination) using the same apparatus as in experiment 1. Participants performed 1 bout of 5 trials for each pattern. Each bout was started with a 10-second visual demonstration of a target coordination pattern at 0.75 Hz, followed by five 20-second trials of replicating the target pattern by moving 1 joystick to move 1 dot on the screen together with another dot controlled by the computer. No visual feedback was provided during each trial.

Data Analysis

PTT was calculated to quantify the eye-hand coordination performance at each coordination pattern. A mixed-design ANOVA was performed on mean PTT data, treating group (participants with a history of concussion vs healthy controls) as a between-subject variable, and target coordination pattern (in-phase vs 90° vs anti-phase) as a within-subject variable. Prior to conducting ANOVAs, the normality of outcome variables was assessed using the Shapiro-Wilk test, which showed that all variables met the normality assumption. Cohen *d* effect size was again calculated to demonstrate the power of the study due to the small sample size.

Results

Experiment 1

Visual Discrimination of Coordination

No group difference was detected for JND at either in-phase or anti-phase (in-phase: $F_{1,10}$=1.11, P=.32; anti-phase, $F_{1,10}$=0.48, P=.50), indicating that the threshold for identifying the in-phase and anti-phase coordination pattern was similar between the participants with a history of concussion ($JND_{in\text{-}phase}$=mean 1.53°, SD 0.46°; $JND_{anti\text{-}phase}$=mean 14.29°, SD 4.87°) and the healthy participants ($JND_{in\text{-}phase}$=mean 1.39°, SD 0.29°; $JND_{anti\text{-}phase}$=mean 13.19°, SD 4.26°). The effect of testing sessions was detected only for JND of anti-phase pattern ($F_{5,50}$=3.01; P=.02), suggesting that both groups increased their ability to discriminate anti-phase coordination patterns in later testing sessions (see Table 1 for test means).

Table . Test means for all task measures across 6 tests (mean, SD) in experiment 1.

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6
JND[a] (degree)						
JND of in-phase	1.66° (0.47°)	1.54° (0.34°)	1.32° (0.36°)	1.30° (0.30°)	1.45° (0.38°)	1.49° (0.42°)
JND of anti-phase	16.80° (4.59°)[b]	14.96° (5.06°)	12.64° (3.57°)	13.80° (5.53°)	11.53° (3.99°)	12.71° (3.09°)
Mean proportion of time-on-task						
Unimanual in-phase	0.61 (0.08)[c]	0.65 (0.13)	0.68 (0.09)	0.71 (0.07)	0.71 (0.09)	0.74 (0.08)
Unimanual anti-phase	0.46 (0.14)	0.50 (0.13)	0.53 (0.19)	0.59 (0.11)	0.56 (0.14)	0.57 (0.11)
Unimanual 90°	0.33 (0.09)[d]	0.43 (0.13)	0.47 (0.13)	0.49 (0.13)	0.52 (0.13)	0.55 (0.14)
Bimanual in-phase	0.70 (0.05)	0.71 (0.06)	0.75 (0.04)	0.74 (0.04)	0.76 (0.04)	0.72 (0.06)
Bimanual anti-phase	0.82 (0.04)	0.84 (0.04)	0.82 (0.04)	0.82 (0.04)	0.81 (0.07)	0.83 (0.04)
Bimanual 90°	0.30 (0.08)[e]	0.37 (0.09)[f]	0.45 (0.06)	0.48 (0.09)	0.45 (0.11)	0.49 (0.12)

[a]JND: just noticeable difference.

[b]Test 1 vs test 5, P=.05.

[c]Test 1 vs test 6, P<.05.

[d]Test 1 vs test 4, P<.05; test 1 vs test 5, P<.01; and test 1 vs test 6, P<.01.

[e]Test 1 vs test 3, P<.01; test 1 vs test 4, P<.01; test 1 vs test 5, P<.01; and test 1 vs test 6, P<.01.

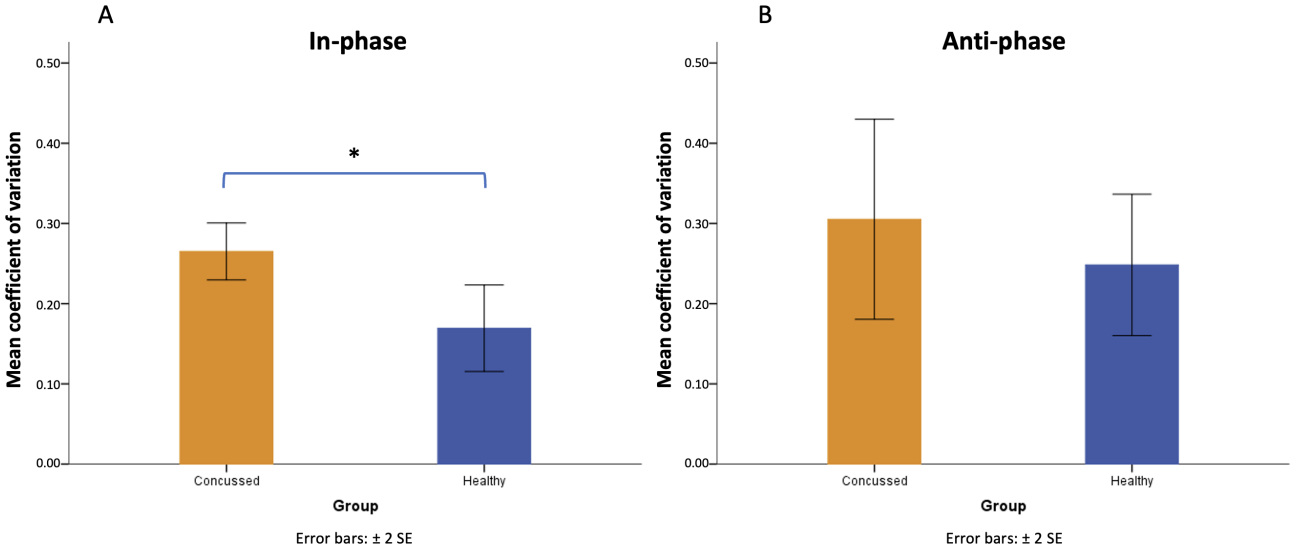
[f]Test 2 vs test 6, P<.05.

The group difference was detected on CV of JND for in-phase coordination pattern (t_{10}=2.93; P=.02) but not for anti-phase coordination pattern (t_{10}=0.77; P=.46). As seen in Figures 2A and 2B, participants in the group with a history of concussion demonstrated a higher inconsistency in their ability to discriminate the in-phase coordination pattern than the healthy

XSL•FO
RenderX

controls across the 6 assessments ($CV_{Concussed}$=mean 0.27, SD 0.04; $CV_{Healthy}$=mean 0.17, SD 0.07), which was not seen in discriminating the anti-phase coordination pattern. No other main effect or interaction was detected (all other *P* values >.05).

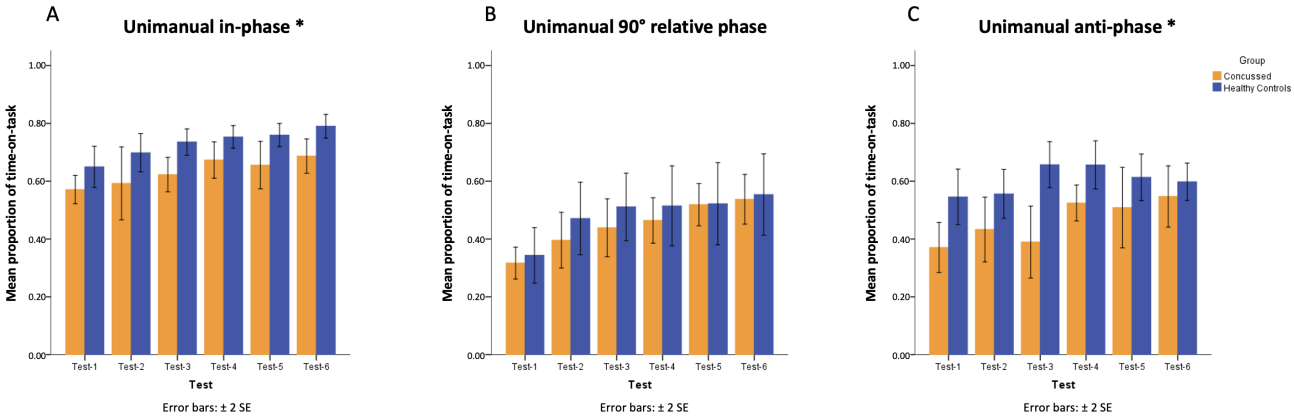
Figure 2. Coefficient of variation of just noticeable difference at in-phase (panel A) and anti-phase pattern (panel B) for participants with a history of concussion and healthy controls. *Significant difference with *P*<.05. Error bars represent the standard error of the mean.



Unimanual Coordination Performance

The main effect of group was detected for performing both in-phase and anti-phase coordination patterns (at in-phase: $F_{1,10}$=8.49, *P*=.02; at anti-phase: $F_{1,10}$=10.67, *P*=.008; Figure 3A-C). Healthy controls overall outperformed participants with a history of concussion in producing the intrinsic coordination patterns (at in-phase: $PTT_{Concussed}$=mean 0.63, SD 0.10, and $PTT_{Healthy}$=mean 0.73, SD 0.08; at anti-phase: $PTT_{Concussed}$=mean 0.46, SD 0.14, and $PTT_{Healthy}$=mean 0.60, SD 0.10). The main effect of testing sessions was detected in performing both intrinsic coordination patterns as well (at in-phase: $F_{5,50}$=7.03, *P*<.001; at anti-phase: $F_{5,50}$=3.16, *P*=.014). In general, healthy controls and participants with a history of concussion both improved their performance of intrinsic coordination patterns in later testing sessions, suggesting that the unimanual coordination task is challenging even for producing the intrinsic coordination patterns (see Table 1 for test means).

Figure 3. Mean proportion of time-on-task at unimanual in-phase (panel A), 90° relative phase (panel B), and anti-phase pattern (panel C) for participants with a history of concussion and healthy controls across 6 tests. *Group main effect revealed by mixed-design ANOVA. Error bars represent the standard error of the mean.



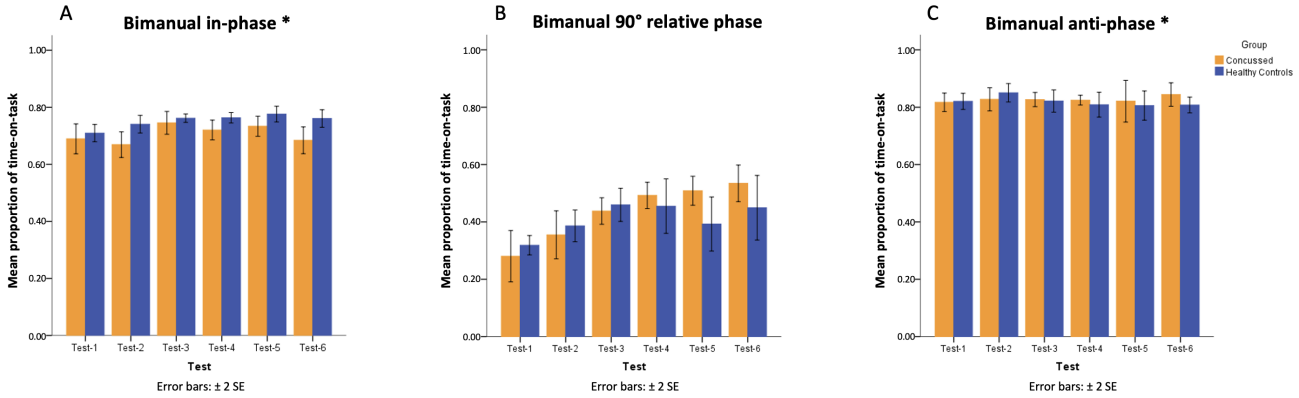
The main effect of testing sessions was detected in performing the novel coordination pattern at the 90° relative phase ($F_{5,50}$=18.92; *P*<.001). Regardless of the history of concussion, all participants learned and improved their unimanual 90° coordination. No other main effect or interaction was detected (all other *P* values >.05).

Bimanual Coordination Performance

The main effect was detected for both group ($F_{1,10}$=6.49; *P*=.03) and testing sessions ($F_{5,50}$=6.15; *P*<.001) in performing the in-phase coordination pattern with no interaction ($F_{5,50}$=1.65; *P*=.16). As seen in Figures 4A-4C, participants with a history of concussion performed the in-phase coordination pattern worse than healthy controls ($PTT_{Concussed}$=mean 0.71, SD 0.06; $PTT_{Healthy}$=mean 0.75, SD 0.04), and both groups improved their performance throughout the testing sessions (Table 1). No main effect or interaction was detected in performing the anti-phase coordination pattern (all other *P* values >.05).

XSL•FO
RenderX

Figure 4. Mean proportion of time-on-task at bimanual in-phase (panel A), 90° relative phase (panel B), and anti-phase pattern (panel C) for participants with a history of concussion and healthy controls across 6 tests. *Group main effect revealed by mixed-design ANOVA. Error bars represent the SE of the mean.



As for the performance of 90° coordination, the significant main effect of testing session ($F_{5,50}$=19.02; P<.001) and the group by testing session interaction ($F_{5,50}$=3.81; P=.005) were detected, but the group difference was insignificant ($F_{1,10}$=0.35; P=.57). Although participants improved their performance at 90° across the testing sessions, their performance changed differently depending on the group. While the health controls dropped their performance in later sessions, participants with a history of concussion maintained their improved performance in later sessions. As shown by the following simple main effect analysis with post hoc of Tukey HSD analysis, participants with a history of concussion outperformed the healthy controls only in the fifth session ($PTT_{Concussed}$=mean 0.51, SD 0.06; $PTT_{Healthy}$=mean 0.39, SD 0.12).

Cohen d Effect Size of All Coordination Tasks

As seen in Table 2, a large effect size (d>0.80) was detected for CV of JND in discriminating the in-phase coordination pattern, PTT in performing the unimanual in-phase and anti-phase coordination tasks, and PTT in performing the bimanual in-phase coordination task, suggesting that these tasks and task measures are promising for screening participants with a history of concussion.

Table . Cohen *d* effect size for all task measures in experiment 1.

Task measures	Main effect		Interaction	Cohen *d* of group main effect
JND[a] of in-phase	N/A[b]	N/A	N/A	0.689
JND of anti-phase	N/A	Test	N/A	0.373
CV[c] of JND at in-phase	Group	N/A	N/A	*1.362* [d]
CV of JND at anti-phase	N/A	N/A	N/A	0.493
PTT[e] of unimanual in-phase	Group	Test	N/A	*1.*591
PTT of unimanual anti-phase	Group	Test	N/A	*1.787*
PTT of unimanual 90°	N/A	Test	N/A	0.314
PTT of bimanual in-phase	Group	Test	N/A	*1.489*
PTT of bimanual anti-phase	N/A	N/A	N/A	0.251
PTT of bimanual 90°	N/A	Test	Group × Test	0.316

[a]JND: just noticeable difference.

[b]N/A: not applicable.

[c]CV: coefficient of variation.

[d]Tasks with over large effect size (Cohen d>0.80) are in italics.

[e]PTT: proportion of time-on-task.

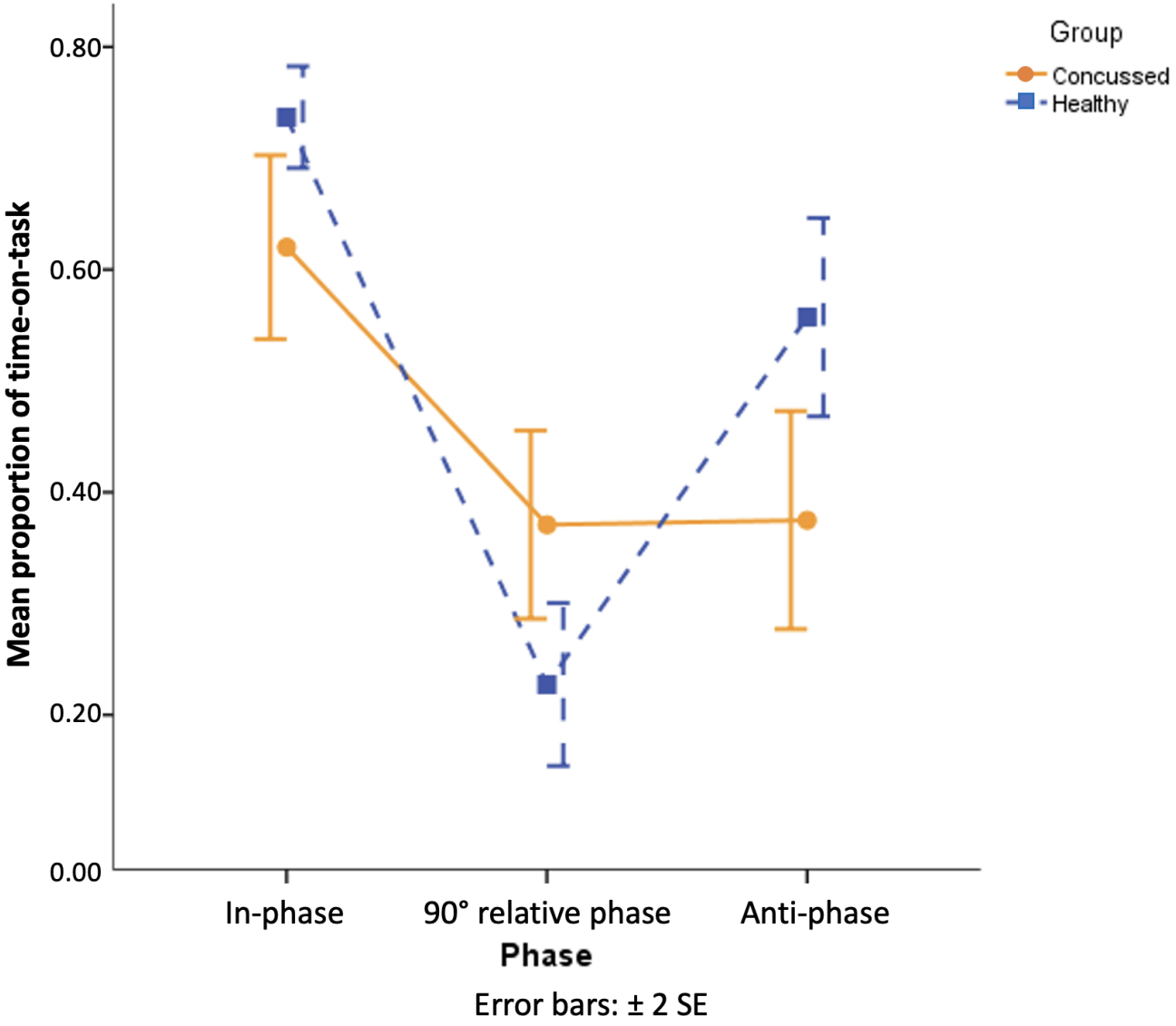
Experiment 2

The 2-way mixed-design ANOVA on mean PTTs yielded a main effect for target coordination pattern ($F_{2,36}$=61.35, P<.001, and Cohen d=3.52). As seen in Figure 5, in general, performance of in-phase coordination (mean PTT 0.68, SD 0.12) was significantly better than that of anti-phase coordination (mean PTT 0.47, SD 0.17) and then that of 90° coordination (mean PTT 0.30, SD 0.14). Significant 2-way interaction between group and target coordination pattern was detected ($F_{2,36}$=12.60, P<.001, and Cohen d=1.54). Simple main effect analysis showed that healthy controls outperformed participants with a history

XSL•FO
RenderX

of concussion in performing in-phase ($F_{1,54}$=4.20, P=.045, and Cohen d=0.48; mean $PTT_{Concussed}$=0.62, SD 0.13; and mean $PTT_{Healthy}$=0.74, SD 0.07) and anti-phase coordination patterns ($F_{1,54}$=10.26, P=.002, and Cohen d=0.81; mean $PTT_{Concussed}$=0.37, SD 0.15; and mean $PTT_{Healthy}$=0.56, SD 0.14), while the participants with a history of concussion outperformed healthy controls only in performing 90° coordination ($F_{1,54}$=6.36, P=.015, and Cohen d=0.62; mean $PTT_{Concussed}$=0.37, SD 0.13; and mean $PTT_{Healthy}$=0.23, SD 0.12). To be noted, the 2 groups demonstrated different patterns in performing the 3 coordination patterns. As revealed by simple main effect analysis followed by post hoc of Tukey HSD test, healthy controls ($F_{2,54}$=41.20, P<.001, and Cohen d=2.38) performed in-phase coordination (mean PTT 0.74, SD 0.07) significantly better than anti-phase coordination (mean PTT 0.56, SD 0.14), while their performance of 90° coordination (mean PTT 0.23, SD 0.12) was significantly worse than that of both in-phase and anti-phase patterns. However, this differentiated performance between the 3 coordination patterns (as healthy controls showed in Figure 5) was absent for participants with a history of concussion ($F_{2,54}$=12.57, P<.001, and Cohen d=1.27) and their performance of in-phase coordination (mean PTT 0.62, SD 0.13) was better than that of anti-phase (mean PTT 0.37, SD 0.15) and 90° coordination (mean PTT 0.37, SD 0.13), with no difference detected between the latter 2, suggesting that their cognitive-motor system could not differentiate the 2 coordination patterns involving displacement of dots.

Figure 5. Mean proportion of time-on-task at unimanual in-phase, unimanual 90° relative phase, and unimanual anti-phase for participants with a history of concussion and healthy controls. A conventional pattern of performance (V-shape) was demonstrated by healthy controls but not by the group with a history of concussion. Error bars represent the SE of the mean.



XSL•FO
RenderX

Discussion

Principal Findings

In experiment 1, three coordination tasks have been implemented to examine the comprehensive cognitive-motor functions of people with and with no history of concussion, and among the 10 task measures, 4 measures have been identified with a great effect size to capture the group difference (Table 2). The results suggest that participants with a history of concussion demonstrated greater variability in perceptual discrimination of the in-phase coordination pattern and worse motor performance in both unimanual and bimanual coordination tasks than healthy controls. These deficits were consistently seen in the following bimonthly sessions, suggesting persistent cognitive-motor impairments in individuals with prior concussion.

Experiment 2 was conducted using an iterated protocol based on the results of experiment 1 and examined the possibility of using the unimanual coordination task alone to detect the cognitive-motor deficits associated with concussion. Compared with the protocol used in experiment 1, the current testing protocol could be completed within 10 minutes (instead of an hour) with the group difference being captured in just 1 testing session. The results showed that unimanual coordination tasks were sensitive to detecting performance deficits in participants with a history of concussion in that individuals with a history of concussion performed worse on both in-phase and anti-phase tasks and failed to differentiate between anti-phase and 90° coordination patterns.

Visual and Motor Coordination Deficits Revealed Through Comprehensive Assessments

In experiment 1, the visual discrimination task was designed to assess cognitive functions of coordination, which include attending to and tracking the 2 moving dots on the screen, memorizing the spatial-temporal pattern of their movements, and then identifying the target coordination pattern from those distracting patterns. Although these cognitive functions have been reported to be impaired by concussion in previous studies [42-44], participants in this study did not show group differences in discriminating the intrinsic coordination patterns, suggesting that our participants with a history of concussion have recovered the cognitive functions required for perceiving the fundamental coordination tasks such as walking and jumping. This is conforming to the literature that most cognitive dysfunctions associated with concussion spontaneously recover to baseline within 2 weeks [45]. Nevertheless, the analysis on the CV of JNDs revealed that our participants with a history of concussion were more inconsistent in discriminating the in-phase pattern than their healthy counterparts, suggesting that the ability to consistently detect the minor decoupling between 2 oscillators (eg, minor asynchronization of foot takeoff in jumping) was impaired following concussion despite the spontaneous recovery of the ability to perceive intrinsic coordination patterns.

The unimanual coordination task was designed to examine cognitive-motor functions of coordination, specifically, the ability to produce and maintain a rhythmic movement to interact with a computer based on the established perception of coordination pattern. Such an ability not only requires the aforementioned cognitive functions to detect and retain the coordination pattern but also demands motor functions to move one hand in coordination with the moving dot on the screen and correct any perceived movement error to maintain the coordination pattern. The group difference was captured when motor functions were demanded together with cognitive functions. Our healthy controls outperformed participants with a history of concussion in performing both intrinsic coordination patterns. Such a finding conforms to a previous study [46] suggesting that the deficits of dynamic motor functions (eg, visuomotor tracking and walking under various conditions of attention) persist following the concussion despite the spontaneous recovery of cognitive functions. As for the performance of the novel coordination pattern, both participants with a history of concussion and healthy participants improved over the testing sessions without showing any group difference, suggesting that the unimanual coordination task was quite challenging in nature, and the ability to learn new coordination pattern remained intact postconcussion.

Finally, the bimanual coordination task was used to examine the motor functions of coordination. Compared with the unimanual task, the bimanual task demands the motor production of the coordination pattern entirely by requiring participants to move 2 hands rhythmically; therefore, the independent visuomotor control and hemisphere cross talk are both challenged. The results showed the group difference only in performing the in-phase pattern with healthy controls outperforming the participants with a history of concussion. Such a finding not only supported previous studies suggesting that bimanual coordination deficits are associated with concussion [47] but also indicated that the concussion-induced malfunction of CC might have caused the problem for spatial and temporal coupling of 2 hands, which is typically seen in patients with split brain or callosotomy [48]. Compared with the anti-phase and novel coordination patterns, the in-phase pattern demands more spatial and temporal coupling of 2 hands. Therefore, people with a history of concussion are more likely to be screened out by the in-phase bimanual coordination task.

In sum, all 3 coordination tasks have been proven effective for discriminating people with a history of concussion with at least 1 task measure that could be used. Interestingly, most effective task measures are related to the ability to perceive or perform the in-phase pattern, suggesting that the cognitive-motor deficits associated with concussions may be specifically tied to the inability to detect and maintain spatiotemporal coupling in performing eye-hand coordination tasks. The unimanual task emerged to be a good candidate for screening people with a history of concussion because (1) half of effective task measures were found in the unimanual task with the largest effect size to detect group difference, (2) it examines the cognitive-motor functions required for both perceiving and performing coordination patterns, and (3) the performance can be quickly assessed and used for evaluation. Consequently, experiment 2 was designed to examine whether the unimanual task alone can serve as a quick and efficient tool for detecting the chronic cognitive-motor deficits following concussion.

XSL•FO
RenderX

Perception-Action Coupling Disruptions in Simplified Unimanual Coordination Task

In experiment 2, the unimanual coordination task entails the integration of cognitive-motor functions including attention, memory, and visuomotor tracking, for which one needs to successfully detect the spatiotemporal relationship between the 2 moving dots on the screen and then adjust the manual movement to maintain that pattern. According to the perception-action theory of coordination [29,49], the relative direction of 2 oscillators is used in perceiving and performing both in-phase and anti-phase coordination patterns, while the relative position of 2 oscillators is used for the 90° coordination. Since in-phase and anti-phase patterns represent the coordination movements with the relative direction always being identical (ie, in-phase) or opposite (ie, anti-phase), they are distinctive and easy to be perceived and produced. In contrast, the relative position is ambiguous in 90° coordination due to the fluctuations in the movement direction, which makes it difficult for both perception and production [50]. Both perceptual and motor studies on rhythmic coordination using healthy participants [31,35,38,51] have shown that the performance of 90° coordination is worse than that of in-phase and anti-phase, with the former outperforming the latter (the differentiated performance between 3 coordination patterns as seen in Figure 5 for healthy controls).

In both experiments when unimanual coordination performance was assessed, we found consistently that healthy controls outperformed participants with a history of concussion in producing both in-phase and anti-phase patterns while maintaining the differentiated performance between all 3 coordination patterns, suggesting that they were sensitive to the information of relative direction that is required for perceiving and performing the intrinsic coordination patterns, although having difficulty in finding the information of relative position for successful perception and production of the novel coordination pattern due to lack of practice. As for the participants with a history of concussion, they showed a significantly degraded performance at both in-phase and anti-phase while losing the performance difference between anti-phase and 90°, suggesting that the previous concussion might have limited their ability to discriminate and use the appropriate spatiotemporal information for successful production of both intrinsic and novel coordination patterns. As shown in experiment 1, the participants with a history of concussion did not have a problem with visual discrimination of in-phase and anti-phase patterns; therefore, their degraded performance in performing unimanual in-phase and anti-phase patterns had to do with the deficit in perception-action coupling. Although their performance at 90° was seemingly better than the healthy controls, such a performance was not different from their performance in producing the anti-phase pattern, suggesting that anti-phase and 90° coordination patterns may have been treated as the same by the individuals with a history of concussion since they both involve the displacement and opposite relative direction of moving oscillators. Hence, their seemingly better performance at 90° was a false positive, stemming from their confusion between 90° coordination and anti-phase coordination.

Although more studies are warranted to prove the impaired ability to discriminate and use spatiotemporal information for coordination by the participants with a history of concussion, experiment 2 overall showed that participants with a history of concussion demonstrated deteriorated cognitive-motor functions in performing the unimanual in-phase and anti-phase patterns, which could be attributed to the dissociation between perception and action. In addition, the conventional differentiated performance between the 3 coordination patterns (in-phase > anti-phase > 90°) was not seen for participants with a history of concussion, which suggests that the differentiated performance between the intrinsic and novel coordination patterns in the unimanual coordination task could potentially serve as a useful indicator for screening concussion.

Limitations and Future Directions

This study has several limitations that should be acknowledged. First, although 2 experiments were conducted, the sample size in each was relatively small, which may limit the generalization of both study findings. While several outcomes demonstrated large effect sizes for group main effect (Cohen d>0.80), which suggest a possibility of using a combined approach for assessment and screening, a larger sample size is warranted in future studies to prove the effectiveness and reliability of using eye-hand coordination tasks to detect the chronic or subacute cognitive-motor deficits associated with concussion.

Second, there was a wide range of the reported time from the last concussion. The reported last concussion occurrence ranged from 2 to 96 months in experiment 1 and from 4 to 54 months in experiment 2. As a result, participants could be at different stages of spontaneous recovery from concussion. Also, during that time, participants could be reinjured by concussion that was not medically examined. Future studies could improve on this by focusing on participants within narrower postinjury time frames or by using recovery stage as a covariate in analyses.

Finally, poor eye-hand coordination could result from neurocognitive disorders other than concussion (eg, posttraumatic stress disorder and attention-deficit/hyperactivity disorder). Although all participants in this study were college students with ongoing study programs, and they self-reported to have no history of substance abuse, peripheral neuropathy, and neurological disorders, there is still a possibility for undiagnosed neurocognitive disorders other than concussion to impact the eye-hand coordination performance. In future work, standardized screening tools or neuropsychological assessments could be included to help isolate the impact of concussion from other potential influences on cognitive-motor performance.

Conclusions

This study provides preliminary evidence for using the computerized eye-hand coordination tasks to detect the chronic cognitive-motor dysfunctions associated with concussion. In developing neurobehavioral tools to capture the postconcussion chronic cognitive-motor deficits, the unimanual coordination task should be considered, and the differentiated performance between intrinsic and novel coordination patterns should be examined.

XSL•FO
RenderX

Acknowledgments

This publication was made possible by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health (grant 2P20GM103432). Publication of this manuscript was made possible by a 2025 Publication Fund award from Texas Woman's University. No generative artificial intelligence tools were used in the writing, data analysis, or interpretation of this manuscript.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

SH was responsible for the study concept and design, data collection, analysis, and drafting of the manuscript. QZ contributed to study design and supervision, resource support, and critical revision of the manuscript. Both authors read and approved the final version of the paper for submission.

Conflicts of Interest

None declared.

References

1. Gardner RC, Yaffe K. Epidemiology of mild traumatic brain injury and neurodegenerative disease. Mol Cell Neurosci 2015 May;66(Pt B):75-80. [doi: 10.1016/j.mcn.2015.03.001] [Medline: 25748121]
2. Jang SH, Seo YS. Diffusion tensor tractography characteristics of axonal injury in concussion/mild traumatic brain injury. Neural Regen Res 2022 May;17(5):978-982. [doi: 10.4103/1673-5374.324825] [Medline: 34558511]
3. Dewan MC, Rattani A, Gupta S, et al. Estimating the global incidence of traumatic brain injury. J Neurosurg 2019 Apr 1;130(4):1080-1097. [doi: 10.3171/2017.10.JNS17352] [Medline: 29701556]
4. Chiang CC, Guo SE, Huang KC, Lee BO, Fan JY. Trajectories and associated factors of quality of life, global outcome, and post-concussion symptoms in the first year following mild traumatic brain injury. Qual Life Res 2016 Aug;25(8):2009-2019. [doi: 10.1007/s11136-015-1215-0] [Medline: 26706751]
5. Fineblit S, Selci E, Loewen H, Ellis M, Russell K. Health-related quality of life after pediatric mild traumatic brain injury/concussion: a systematic review. J Neurotrauma 2016 Sep 1;33(17):1561-1568. [doi: 10.1089/neu.2015.4292] [Medline: 26916876]
6. Kozlowski Moreau O, Blin P, Guillemot M, Daveluy W, Deicas P. Communication, social skills and quality of life after mild traumatic brain injury: a comparison of subjects with or without complaints. Ann Phys Rehabil Med 2015 Sep;58:e148. [doi: 10.1016/j.rehab.2015.07.352]
7. Cassidy JD, Cancelliere C, Carroll LJ, et al. Systematic review of self-reported prognosis in adults after mild traumatic brain injury: results of the International Collaboration on Mild Traumatic Brain Injury Prognosis. Arch Phys Med Rehabil 2014 Mar;95(Suppl 3):S132-S151. [doi: 10.1016/j.apmr.2013.08.299] [Medline: 24581902]
8. Helmick K, Baugh L, Lattimore T, Goldman S. Traumatic brain injury: next steps, research needed, and priority focus areas. Mil Med 2012 Aug;177(8 Suppl):86-92. [doi: 10.7205/milmed-d-12-00174] [Medline: 22953445]
9. Traumatic brain injury: hope through research. NINDS and NIH. 2020. URL: https://catalog.ninds.nih.gov/sites/default/files/publications/traumatic-brain-injury-hope-through-research_1.pdf [accessed 2025-07-30]
10. Bernier RA, Hillary FG. Traumatic brain injury and frontal lobe plasticity. Handb Clin Neurol 2019;163:411-431. [doi: 10.1016/B978-0-12-804281-6.00022-7] [Medline: 31590744]
11. Francis HM, Osborne-Crowley K, McDonald S. Validity and reliability of a questionnaire to assess social skills in traumatic brain injury: a preliminary study. Brain Inj 2017;31(3):336-343. [doi: 10.1080/02699052.2016.1250954] [Medline: 28102714]
12. Miller KJ, Ivins BJ, Schwab KA. Self-reported mild TBI and postconcussive symptoms in a peacetime active duty military population: effect of multiple TBI history versus single mild TBI. J Head Trauma Rehabil 2013;28(1):31-38. [doi: 10.1097/HTR.0b013e318255ceae] [Medline: 22647963]
13. Vynorius KC, Paquin AM, Seichepine DR. Lifetime multiple mild traumatic brain injuries are associated with cognitive and mood symptoms in young healthy college students. Front Neurol 2016;7:188. [doi: 10.3389/fneur.2016.00188] [Medline: 27843436]
14. Misquitta K, Dadar M, Tarazi A, et al. The relationship between brain atrophy and cognitive-behavioural symptoms in retired Canadian football players with multiple concussions. Neuroimage Clin 2018;19:551-558. [doi: 10.1016/j.nicl.2018.05.014] [Medline: 29984163]
15. Ptito A, Chen JK, Johnston KM. Contributions of functional magnetic resonance imaging (fMRI) to sport concussion evaluation. Neurorehabilitation 2007;22(3):217-227. [Medline: 17917172]

XSL•FO
RenderX

16. van der Horn HJ, de Haan S, Spikman JM, de Groot JC, van der Naalt J. Clinical relevance of microhemorrhagic lesions in subacute mild traumatic brain injury. Brain Imaging Behav 2018 Jun;12(3):912-916. [doi: 10.1007/s11682-017-9743-6] [Medline: 28664231]
17. McInnes K, Friesen CL, MacKenzie DE, Westwood DA, Boe SG. Mild traumatic brain injury (mTBI) and chronic cognitive impairment: a scoping review. PLoS One 2017;12(4):e0174847. [doi: 10.1371/journal.pone.0174847]
18. McCrory P, Meeuwisse W, Dvořák J, et al. Consensus statement on concussion in sport—the 5th international conference on concussion in sport held in Berlin, October 2016. Br J Sports Med 2017 Jun;51(11):838-847. [doi: 10.1136/bjsports-2017-097699] [Medline: 28446457]
19. Abe M, Hanakawa T. Functional coupling underlying motor and cognitive functions of the dorsal premotor cortex. Behav Brain Res 2009 Mar 2;198(1):13-23. [doi: 10.1016/j.bbr.2008.10.046] [Medline: 19061921]
20. Gooijers J, Beets IAM, Albouy G, et al. Movement preparation and execution: differential functional activation patterns after traumatic brain injury. Brain (Bacau) 2016 Sep;139(Pt 9):2469-2485. [doi: 10.1093/brain/aww177] [Medline: 27435093]
21. McAfee SS, Liu Y, Sillitoe RV, Heck DH. Cerebellar coordination of neuronal communication in cerebral cortex. Front Syst Neurosci 2021;15:781527. [doi: 10.3389/fnsys.2021.781527] [Medline: 35087384]
22. Leonard HC, Hill EL. Executive difficulties in developmental coordination disorder: methodological issues and future directions. Curr Dev Disord Rep 2015 Jun;2(2):141-149. [doi: 10.1007/s40474-015-0044-8]
23. Schneiders AG, Sullivan SJ, Gray AR, Hammond-Tooke GD, McCrory PR. Normative values for three clinical measures of motor performance used in the neurological assessment of sports concussion. J Sci Med Sport 2010 Mar;13(2):196-201. [doi: 10.1016/j.jsams.2009.05.004] [Medline: 19560971]
24. Kelty-Stephen DG, Qureshi Ahmad M, Stirling L. Use of a tracing task to assess visuomotor performance for evidence of concussion and recuperation. Psychol Assess 2015 Dec;27(4):1379-1387. [doi: 10.1037/pas0000122] [Medline: 25894704]
25. Chen HL, Lu TW, Chou LS. Effect of concussion on inter-joint coordination during divided-attention gait. J Med Biol Eng 2015 Feb;35(1):28-33. [doi: 10.1007/s40846-015-0002-2]
26. De Beaumont L, Mongeon D, Tremblay S, et al. Persistent motor system abnormalities in formerly concussed athletes. J Athl Train 2011;46(3):234-240. [doi: 10.4085/1062-6050-46.3.234] [Medline: 21669091]
27. Chamard E, Lefebvre G, Lassonde M, Theoret H. Long-term abnormalities in the corpus callosum of female concussed athletes. J Neurotrauma 2016 Jul 1;33(13):1220-1226. [doi: 10.1089/neu.2015.3948] [Medline: 26414496]
28. Ting WKC, Schweizer TA, Topolovec-Vranic J, Cusimano MD. Antisaccadic eye movements are correlated with corpus callosum white matter mean diffusivity, stroop performance, and symptom burden in mild traumatic brain injury and concussion. Front Neurol 2015;6:271. [doi: 10.3389/fneur.2015.00271] [Medline: 26834693]
29. Snapp-Childs W, Wilson AD, Bingham GP. The stability of rhythmic movement coordination depends on relative speed: the Bingham model supported. Exp Brain Res 2011 Nov;215(2):89-100. [doi: 10.1007/s00221-011-2874-x] [Medline: 21952789]
30. Huang S, Layer J, Smith D, Bingham GP, Zhu Q. The effect of movement frequency on perceptual-motor learning of a novel bimanual coordination pattern. Hum Mov Sci 2022 Jun;83:102958. [doi: 10.1016/j.humov.2022.102958] [Medline: 35561528]
31. Snapp-Childs W, Wilson AD, Bingham GP. Transfer of learning between unimanual and bimanual rhythmic movement coordination: transfer is a function of the task dynamic. Exp Brain Res 2015 Jul;233(7):2225-2238. [doi: 10.1007/s00221-015-4292-y] [Medline: 25929551]
32. Zhu Q, Mirich T, Huang S, Snapp-Childs W, Bingham GP. When kinesthetic information is neglected in learning a novel bimanual rhythmic coordination. Atten Percept Psychophys 2017 Aug;79(6):1830-1840. [doi: 10.3758/s13414-017-1336-3] [Medline: 28508115]
33. Huang S, Dai B, Zhu Q. Advantage of early focus on visual information in bi-modal training of bimanual coordination. Multisens Res 2019 Jan 1;1:1-21. [doi: 10.1163/22134808-20191424] [Medline: 31247584]
34. Wilson AD, Tresilian J, Schlaghecken F. The masked priming toolbox: an open-source MATLAB toolbox for masked priming researchers. Behav Res Methods 2011 Mar;43(1):210-214. [doi: 10.3758/s13428-010-0034-z] [Medline: 21287113]
35. Kelso JA. Phase transitions and critical behavior in human bimanual coordination. Am J Physiol 1984 Jun;246(6 Pt 2):R1000-R1004. [doi: 10.1152/ajpregu.1984.246.6.R1000] [Medline: 6742155]
36. Scholz JP, Kelso JA. Intentional switching between patterns of bimanual coordination depends on the intrinsic dynamics of the patterns. J Mot Behav 1990 Mar;22(1):98-124. [doi: 10.1080/00222895.1990.10735504] [Medline: 15111283]
37. Zanone PG, Kelso JA. Evolution of behavioral attractors with learning: nonequilibrium phase transitions. J Exp Psychol Hum Percept Perform 1992 May;18(2):403-421. [doi: 10.1037//0096-1523.18.2.403] [Medline: 1593227]
38. Zaal FT, Bingham GP, Schmidt RC. Visual perception of mean relative phase and phase variability. J Exp Psychol Hum Percept Perform 2000 Jun;26(3):1209-1220. [doi: 10.1037//0096-1523.26.3.1209] [Medline: 10884018]
39. Kingdom FAA, Prins N. Psychophysics: A Practical Introduction: Academic Press; 2010.
40. Huang S, Van Syoc B, Yang R, Kuehn T, Smith D, Zhu Q. Using visual and/or kinesthetic information to stabilize intrinsic bimanual coordination patterns is a function of movement frequency. Psychol Res 2021 Mar;85(2):865-878. [doi: 10.1007/s00426-020-01288-2] [Medline: 31989241]
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences: Academic press; 2013.

XSL•FO
RenderX

42. Moore DR, Pindus DM, Raine LB, et al. The persistent influence of concussion on attention, executive control and neuroelectric function in preadolescent children. Int J Psychophysiol 2016 Jan;99:85-95. [doi: 10.1016/j.ijpsycho.2015.11.010] [Medline: 26608697]
43. Shah-Basak PP, Urbain C, Wong S, et al. Concussion alters the functional brain processes of visual attention and working memory. J Neurotrauma 2018 Jan 15;35(2):267-277. [doi: 10.1089/neu.2017.5117] [Medline: 29020848]
44. Tapper A, Gonzalez D, Roy E, Niechwiej-Szwedo E. Executive function deficits in team sport athletes with a history of concussion revealed by a visual-auditory dual task paradigm. J Sports Sci 2017 Feb;35(3):231-240. [doi: 10.1080/02640414.2016.1161214] [Medline: 26999459]
45. Iverson GL, Gardner AJ, Terry DP, et al. Predictors of clinical recovery from concussion: a systematic review. Br J Sports Med 2017 Jun;51(12):941-948. [doi: 10.1136/bjsports-2017-097729] [Medline: 28566342]
46. Parker TM, Osternig LR, van Donkelaar P, Chou LS. Recovery of cognitive and dynamic motor function following concussion. Br J Sports Med 2007 Dec;41(12):868-873. [doi: 10.1136/bjsm.2006.033761] [Medline: 17517857]
47. Dyck A. Increased GFAP in Motor Cortex and Bimanual Coordination Deficits in a Rat (Rattus norvegicus) Model of Repeated Pediatric Mild Traumatic Brain Injury [master's thesis]. : University of Manitoba; 2018 URL: https://mspace.lib.umanitoba.ca/xmlui/handle/1993/33356 [accessed 2025-07-20]
48. Gooijers J, Swinnen SP. Interactions between brain structure and behavior: the corpus callosum and bimanual coordination. Neurosci Biobehav Rev 2014 Jun;43:1-19. [doi: 10.1016/j.neubiorev.2014.03.008] [Medline: 24661987]
49. Bingham GP, Snapp-Childs W, Zhu Q. Information about relative phase in bimanual coordination is modality specific (not amodal), but kinesthesis and vision can teach one another. Hum Mov Sci 2018 Aug;60:98-106. [doi: 10.1016/j.humov.2018.05.010] [Medline: 29852338]
50. Wilson AD, Bingham GP. Identifying the information for the visual perception of relative phase. Percept Psychophys 2008 Apr;70(3):465-476. [doi: 10.3758/pp.70.3.465] [Medline: 18459257]
51. Bingham GP, Zaal FT, Shull JA, Collins DR. The effect of frequency on the visual perception of relative phase and phase variability of two oscillating objects. Exp Brain Res 2001 Feb;136(4):543-552. [doi: 10.1007/s002210000610] [Medline: 11291735]

Abbreviations

CC: corpus callosum
CV: coefficient of variation
HSD: Honestly Significant Difference
JND: just noticeable difference
mTBI: mild traumatic brain injury
PTT: proportion of time-on-task

Edited by TDA Cardoso; submitted 04.01.25; peer-reviewed by H Tsukamoto, X Wang; revised version received 14.06.25; accepted 16.06.25; published 05.08.25.

Please cite as:
Zhu Q, Huang S
Detection of Chronic Cognitive-Motor Deficits in Adults With a History of Concussion Using Computerized Eye-Hand Coordination Tasks: Preliminary Experimental Design Study
Interact J Med Res 2025;14:e70867
URL: https://www.i-jmr.org/2025/1/e70867
doi:10.2196/70867

© Qin Zhu, Shaochen Huang. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 5.8.2025. 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 Factor Trends of Top Obstetrics and Gynecology Journals During COVID-19

Minhazur Sarker[1], MD; Emily Yang[2], MD; Ukachi Emeruwa[1], MPH, MD; Timothy Wen[1,3], MPH, MD

[1]Division of Maternal Fetal Medicine, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Diego, 9300 Campus Point Dr, Mail Code 7433, San Diego, CA, United States

[2]Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Diego, San Diego, CA, United States

[3]Department of Medicine, University of California, San Diego, San Diego, CA, United States

Corresponding Author:
Minhazur Sarker, MD
Division of Maternal Fetal Medicine, Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Diego, 9300 Campus Point Dr, Mail Code 7433, San Diego, CA, United States

Abstract

Obstetrics and gynecology journal impact factor trends during the COVID-19 pandemic were similar to those seen among other medical specialties, and our findings further highlight the ongoing need to implement a metric of research impact that is not as easily manipulated by selective publication.

(Interact J Med Res 2025;14:e70554) doi:10.2196/70554

KEYWORDS

impact factor; COVID-19; research impact

Introduction

The impact factor (IF), originally introduced as a metric to aid librarian purchasing decisions, has quickly become a marker for journal prestige [1]. Over the years, the concept of IFs has been highly criticized, as they are susceptible to manipulation and may be a disservice to science and their readers [1-4]. While temporal changes may provide insight into journal growth, IFs are also uniquely affected by publication trends. Studies have shown that the COVID-19 pandemic created an unprecedented surge of highly impactful publications increasing IFs [5-8]. Whether an IF uptrend was appreciated among obstetrics and gynecology (OBGYN) journals during COVID-19 and whether the changes reflect a COVID-19 "blockbuster" effect or genuine advancement in women's health research remains unknown. We aimed to determine the IF trends of the top OBGYN journals and compare them to representative journals for other medical specialties (non-OBGYN).

Methods

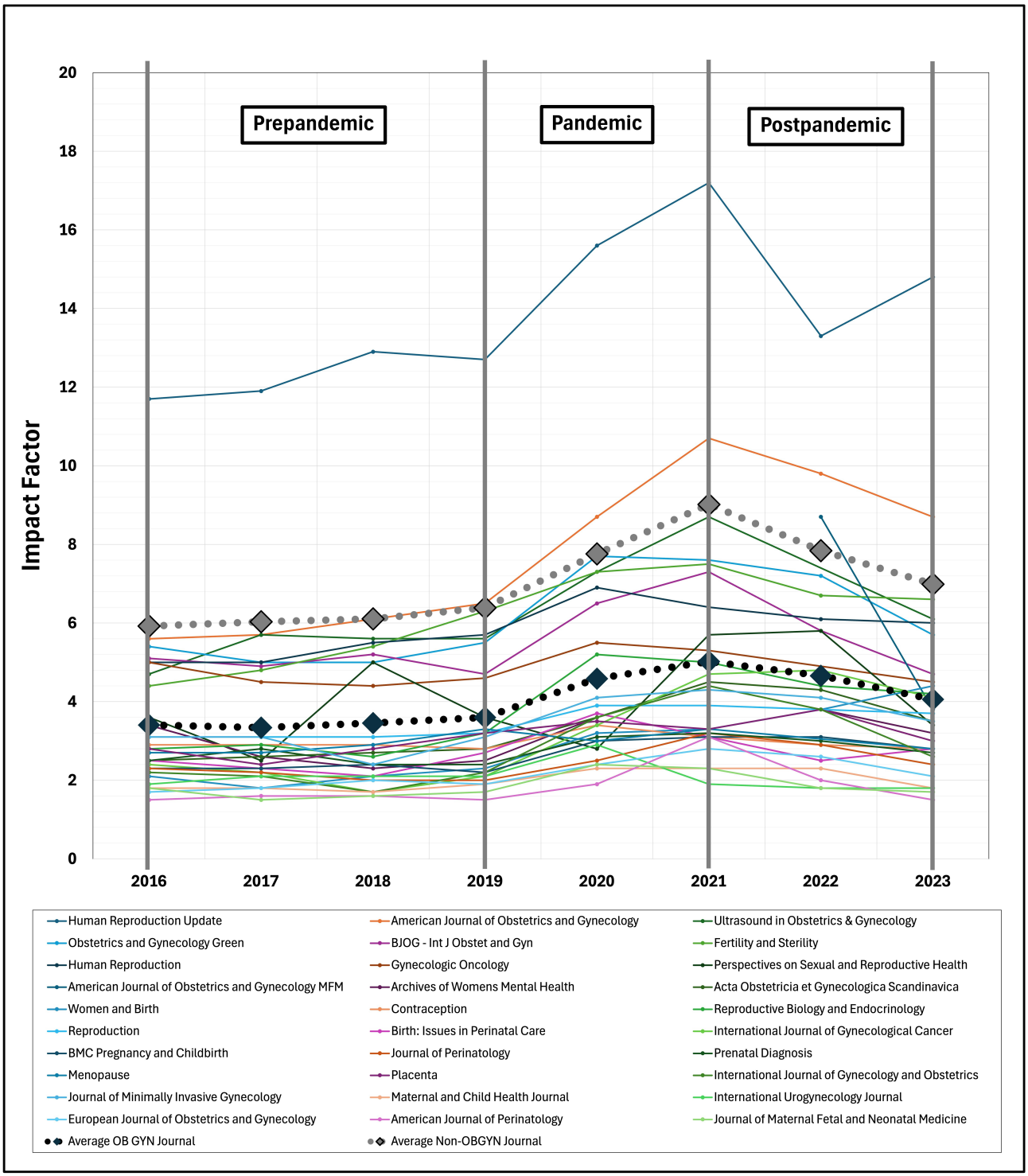
Using the SCImago Journal Rank database, we identified the top 30 broad scope women's health journals (Multimedia Appendix 1). Non-OBGYN national societies were queried to highlight each subspecialty's representative journal. Annual IFs were identified using a web-based searchable database, and temporal trends were assessed over three time periods: (1) prepandemic (2016 - 2019), (2) COVID-19 pandemic (2019 - 2021), and (3) postpandemic (2021 - 2023) [9]. IF trends by individual journal and composite journal type were assessed using the National Cancer Institute's Joinpoint Regression Program, with measures of association expressed with annual average percent changes (AAPCs) and 95% CIs [10]. The average IF percent changes between composite journal types were compared using Student *t* test. Institutional review board approval was not required for this study.

Results

IF data obtained from the top 30 OBGYN journals were compared to 19 non-OBGYN journals. From 2016 to 2023, the mean OBGYN journal IF increased from 3.4 to 4.1 (AAPC 4.1%, 95% CI 0.8%-4.1%) compared to 5.9 to 7.0 (AAPC 3.6%, 95% CI 0.4%-7.5%) for non-OBGYN journals with an average percent change of 18.8% compared to 16.3%, respectively (*P*=.91; Figures 1 and 2).

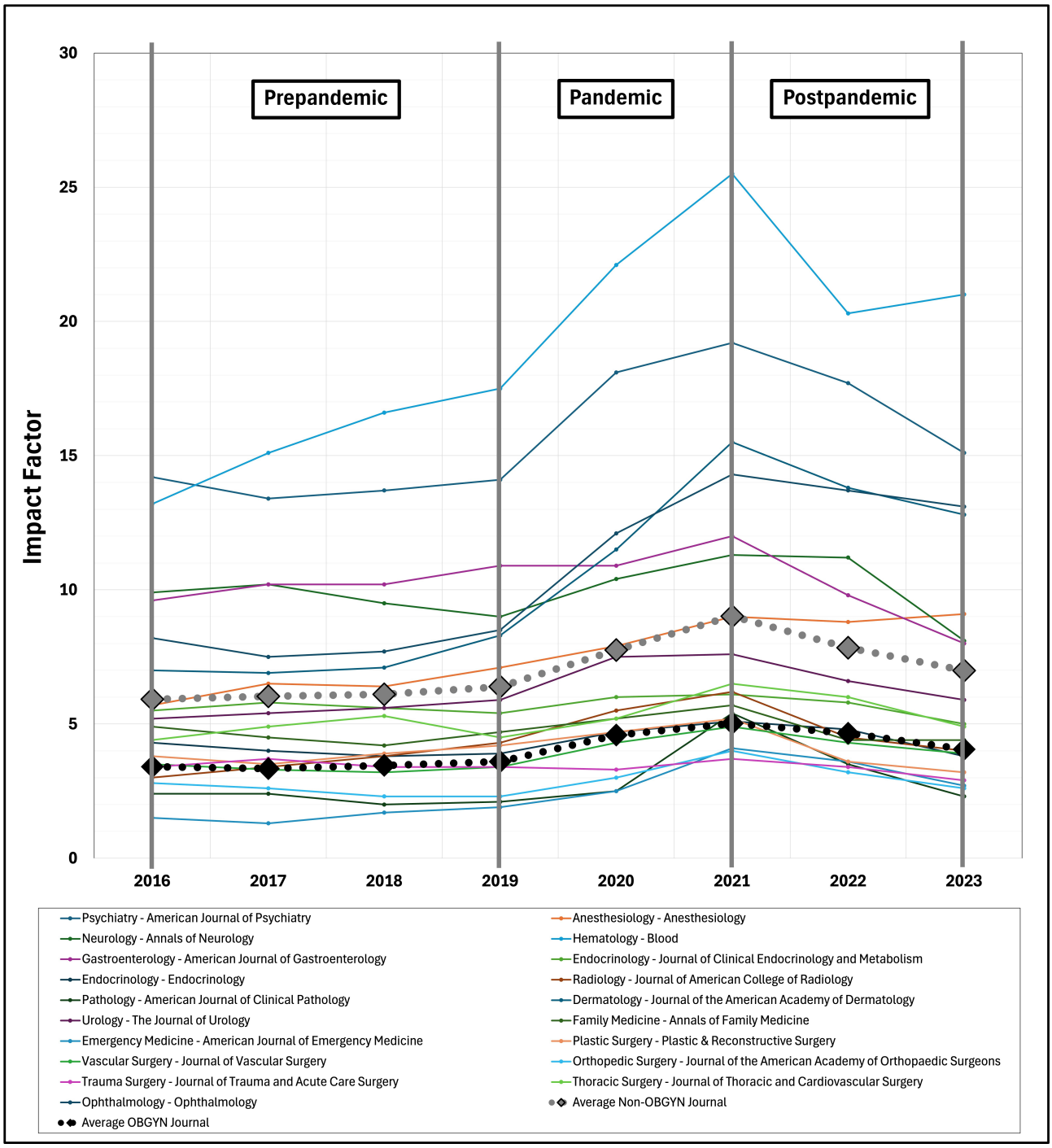
XSL•FO
RenderX

Figure 1. Impact factor trends among the top 30 OBGYN journals. The dotted black line represents the average impact factor among OBGYN journals, while the dotted gray line represents the average impact factor among non-OBGYN journals. OBGYN: obstetrics and gynecology.



XSL•FO
RenderX

Figure 2. Impact factor trends among 19 non-OBGYN medical specialty-specific journals. The dotted black line represents the average impact factor among OBGYN journals, while the dotted gray line represents the average IF among non-OBGYN journals. OBGYN: obstetrics and gynecology.



In the prepandemic period, the mean OBGYN journal IF increased from 3.4 to 3.6 (AAPC 8.9%, 95% CI 4.7%-18.1%) compared to 5.9 to 6.4 (AAPC 8.8%, 95% CI 5.0%-17.8%) for non-OBGYN journals with an average percent change of 4.4% compared to 6.9%, respectively (*P*=.57). During the pandemic, the mean OBGYN journal IF increased from 3.6 to 5.0 (AAPC 8.9%, 95% CI –1.1% to 13.6%) compared to 6.4 to 9.0 (AAPC 8.8%, 95% CI –1.8% to 13.4%) for non-OBGYN journals with an average percent change of 42.7% compared to 47.6%, respectively (*P*=.63). Finally, in the postpandemic period, the mean OBGYN journal IF decreased from 5.0 to 4.1 (AAPC –7.2%, 95% CI –18.2% to 4.3%) compared to 9.0 to 7.0 (AAPC –8.3%, 95% CI –19.0% to 2.9%) for non-OBGYN journals with an average percent change of –18.3% compared to –25.7%, respectively (*P*=.16).

Discussion

Our findings indicated significant increases in overall IF trends from 2016 to 2023 among OBGYN journals, though not statistically different from non-OBGYN journals. When analyzing prepandemic, during the pandemic, and postpandemic periods, both OBGYN and non-OBGYN journals exhibited a similar pattern of small increases prepandemic, robust uptrends during the pandemic, then a steep but less pronounced decline post pandemic (Figures 1 and 2). Given these results, it is likely

XSL•FO
RenderX

that IF uptrends were an artificial COVID-19–related "blockbuster" effect rather than genuine advancements in women's health research.

The COVID-19 trends shown here among OBGYN journals have similarly been seen among medicine, global health, and infectious disease journals during the pandemic [5,6,8]. On one hand, the rapid surge in publications shows how strongly we can mobilize and disseminate research early and effectively to quickly improve health outcomes. On the other hand, there was pressure among journals to publish COVID-19 literature given its highly impactful and citable nature. Beyond being used for an indication it was not intended for, the IF has now repeatedly been highly criticized by many studies yet remains the most mainstream metric to discuss journal importance or prestige.

Our study has limitations. We are unable to comment on authors' research funding changes or each journal's editorial policies and shifts in publication practices during the pandemic. Additionally, our analysis was unable to adjust for journal size, scope, or publication frequency.

Using IFs alone in their current capacity may lead to wrongly concluding genuine research advancement in OBGYN. Among the public, this misunderstanding of the IF has the potential to create further distrust of medical literature. Our study further highlights the ongoing need to implement a different metric of research impact such as a composite using the IF, the *h*-index, and altmetrics that are not as easily manipulated by selective publication.

Acknowledgments

MS's time is supported by a National Institutes of Health T32 HD007203 grant. Funding bodies had no input in study design or interpretation. This study is not a clinical trial. This work has never been presented. Generative artificial intelligence was not used for any portion of this manuscript.

Data Availability

The data used for this study are publicly available as referenced in our text.

Authors' Contributions

Conceptualization: MS, TW
Formal analysis: MS, TW
Investigation: MS
Methodology: MS, EY, UE, TW
Writing – original draft: MS
Writing – review and editing: MS, EY, UE, TW
Supervision: TW

Conflicts of Interest

TW serves on the medical advisory board for Delfina Care, Inc and retains equity and a stipend as compensation. The work presented here is not related to his work with Delfina Care, Inc. No other disclosures have been reported by the authors.

Multimedia Appendix 1
SCImago Journal Rank search strategy.
[DOCX File, 777 KB - i-jmr_v14i1e70554_app1.docx]

References

1. Jacobs K. The impact factor: the ubiquitous metric in evaluating scientific journals. Work 2024;79(1):1-2. [doi: 10.3233/WOR-246016] [Medline: 39213124]
2. Chew M, Villanueva EV, Van Der Weyden MB. Life and times of the impact factor: retrospective analysis of trends for seven medical journals (1994-2005) and their editors' views. J R Soc Med 2007 Mar;100(3):142-150. [doi: 10.1177/014107680710000313] [Medline: 17339310]
3. Falagas ME, Zouglakis GM, Papastamataki PA. Trends in the impact factor of scientific journals. Mayo Clin Proc 2006 Oct;81(10):1401-1402. [doi: 10.4065/81.10.1401] [Medline: 17036568]
4. Brown H. How impact factors changed medical publishing--and science. BMJ 2007 Mar 17;334(7593):561-564. [doi: 10.1136/bmj.39142.454086.AD] [Medline: 17363824]
5. Delardas O, Giannos P. How COVID-19 affected the journal impact factor of high impact medical journals: bibliometric analysis. J Med Internet Res 2022 Dec 21;24(12):e43089. [doi: 10.2196/43089] [Medline: 36454727]
6. He J, Liu X, Lu X, et al. The impact of COVID-19 on global health journals: an analysis of impact factor and publication trends. BMJ Glob Health 2023 Apr;8(4):e011514. [doi: 10.1136/bmjgh-2022-011514] [Medline: 37012001]

XSL•FO
RenderX

7. Uddin S, Khan A, Lu H. Impact of COVID-19 on journal impact factor. J Informetr 2023 Nov;17(4):101458. [doi: 10.1016/j.joi.2023.101458]
8. Maillard A, Delory T. Blockbuster effect of COVID-19 on the impact factor of infectious disease journals. Clin Microbiol Infect 2022 Dec;28(12):1536-1538. [doi: 10.1016/j.cmi.2022.08.011] [Medline: 35998808]
9. Bioxbio. URL: https://www.bioxbio.com/ [accessed 2025-06-06]
10. Joinpoint Trend Analysis Software. National Cancer Institute. URL: https://surveillance.cancer.gov/joinpoint/ [accessed 2025-06-06]

Abbreviations

AAPC: annual average percent change
IF: impact factor
OBGYN: obstetrics and gynecology

Edited by TDA Cardoso; submitted 24.12.24; peer-reviewed by E Gavor, S Uddin; revised version received 12.05.25; accepted 14.05.25; published 17.06.25.

Please cite as:
Sarker M, Yang E, Emeruwa U, Wen T
Impact Factor Trends of Top Obstetrics and Gynecology Journals During COVID-19
Interact J Med Res 2025;14:e70554
URL: https://www.i-jmr.org/2025/1/e70554
doi:10.2196/70554

© Minhazur Sarker, Emily Yang, Ukachi Emeruwa, Timothy Wen. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 17.6.2025. 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

Population Estimates of Self-Reported Depression and Anxiety in the US From a National Survey: Cross-Sectional Survey Study

Yuvraj Pathak, PhD; Elvira Makk-Frid, MPP

West Health, 10350 N Torrey Pines Rd, San Diego, CA, United States

Corresponding Author:
Yuvraj Pathak, PhD
West Health, 10350 N Torrey Pines Rd, San Diego, CA, United States

Abstract

This letter shows that an estimated 31 million and 44 million adults self-report near-daily experiences of depression or anxiety, respectively. Of these, nearly a third have never spoken to a health care provider about it.

(Interact J Med Res 2025;14:e70626) doi:10.2196/70626

KEYWORDS

depression; anxiety; survey

Introduction

Mental health conditions, including depression and anxiety, have been on the rise in the United States over the past several years [1]. Data from the 2022 National Health Interview Survey indicate that the percentage of adults with anxiety and depression symptoms increased from 15.6% to 18.2% and 18.5% to 21.4%, respectively, between 2019 and 2022 [2]. Depression and anxiety can manifest as fatigue, difficulty concentrating, problems sleeping, or changes in appetite and negatively affect daily life [3,4].

Despite these impacts, people may not talk to their health care provider (HCP) due to stigma, cost, lack of access, or lack of insurance coverage [5]. In particular, stereotypes and prejudice can prevent people from discussing their mental health experiences with friends, family, or even HCPs, which could prevent them from receiving needed care [6].

In this paper, we estimate the number of people experiencing depression and anxiety almost daily based on a recent nationally representative survey. We also estimate how many of these people have never spoken to an HCP about their mental health. To the best of our knowledge, we are the first to report population-level estimates of self-reported experiences of depression and anxiety. While these results do not correspond to clinical diagnoses or a definite need for treatment, they provide timely information on an important issue and highlight the need to address mental and emotional health.

Methods

Ethical Considerations

The data for this cross-sectional study were obtained from a nationally representative West Health–Gallup web and mail survey conducted from November 2023 to January 2024; 5149 adults aged ≥18 years were sampled from the Gallup panel using stratified sampling spanning all 50 states and the District of Columbia. The survey response rate was 38%. Ethical approval was granted by Gallup's internal institutional review board. Participation was voluntary, and responses were deidentified. No personal health information or other sensitive information was disclosed to the authors. Multimedia Appendix 1 provides details.

Statistics

Percentages were estimated as the count of responses divided by the number of respondents, multiplied by 100. Probability sampling weights provided by Gallup were used to calculate the population estimates.

Outcomes

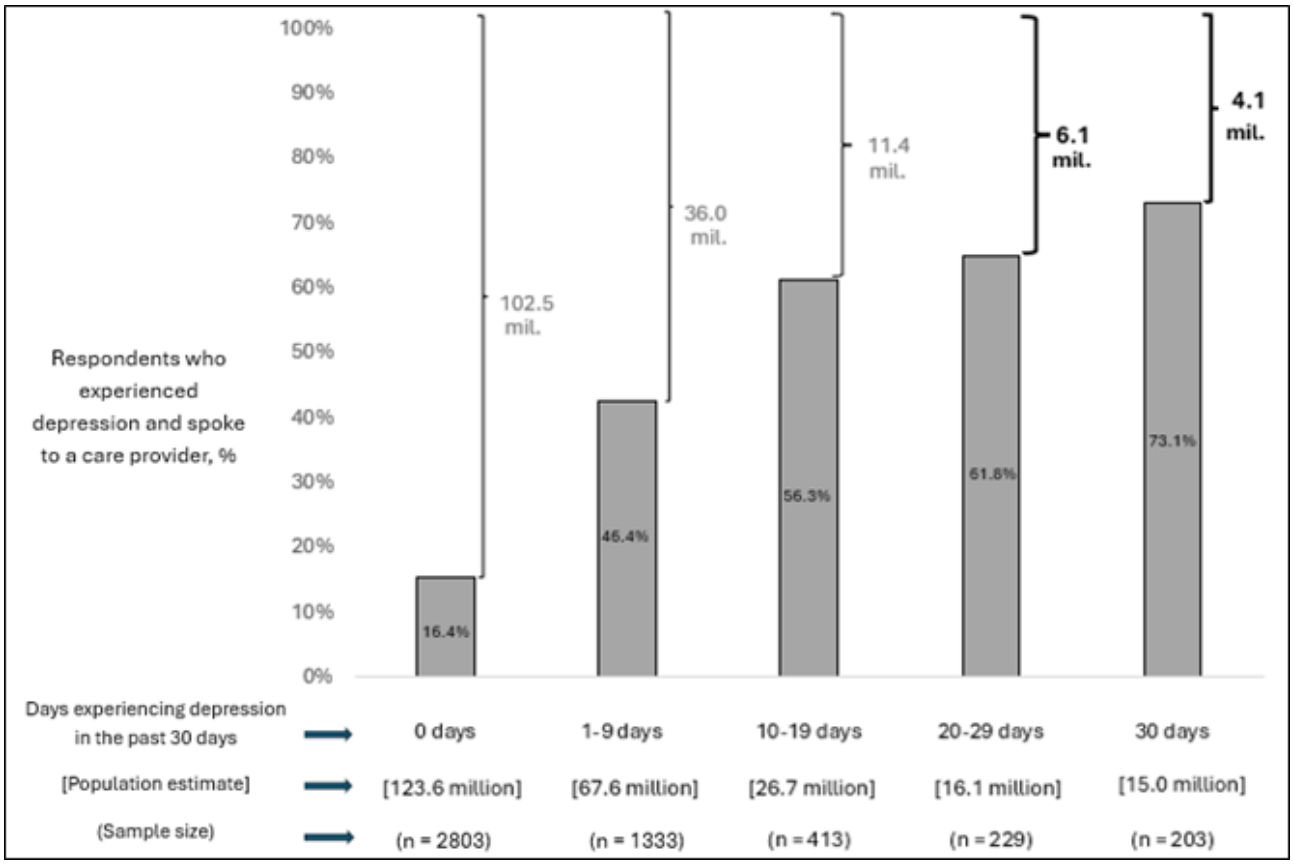
Survey questions focused on perceptions about health care. Our 2 main outcomes were self-reported experiences of depression and anxiety. The survey asked respondents how many days they experienced either condition in the past 30 days and if they had ever discussed depression or anxiety with their HCP. Multimedia Appendix 1 provides more details.

Results

We estimate that 31.1 million US adults experienced symptoms of depression on 20 or more days in the past month at the time of this survey. Of these people, an estimated 10.2 million never spoke to an HCP about their mental health (Figure 1).

XSL•FO
RenderX

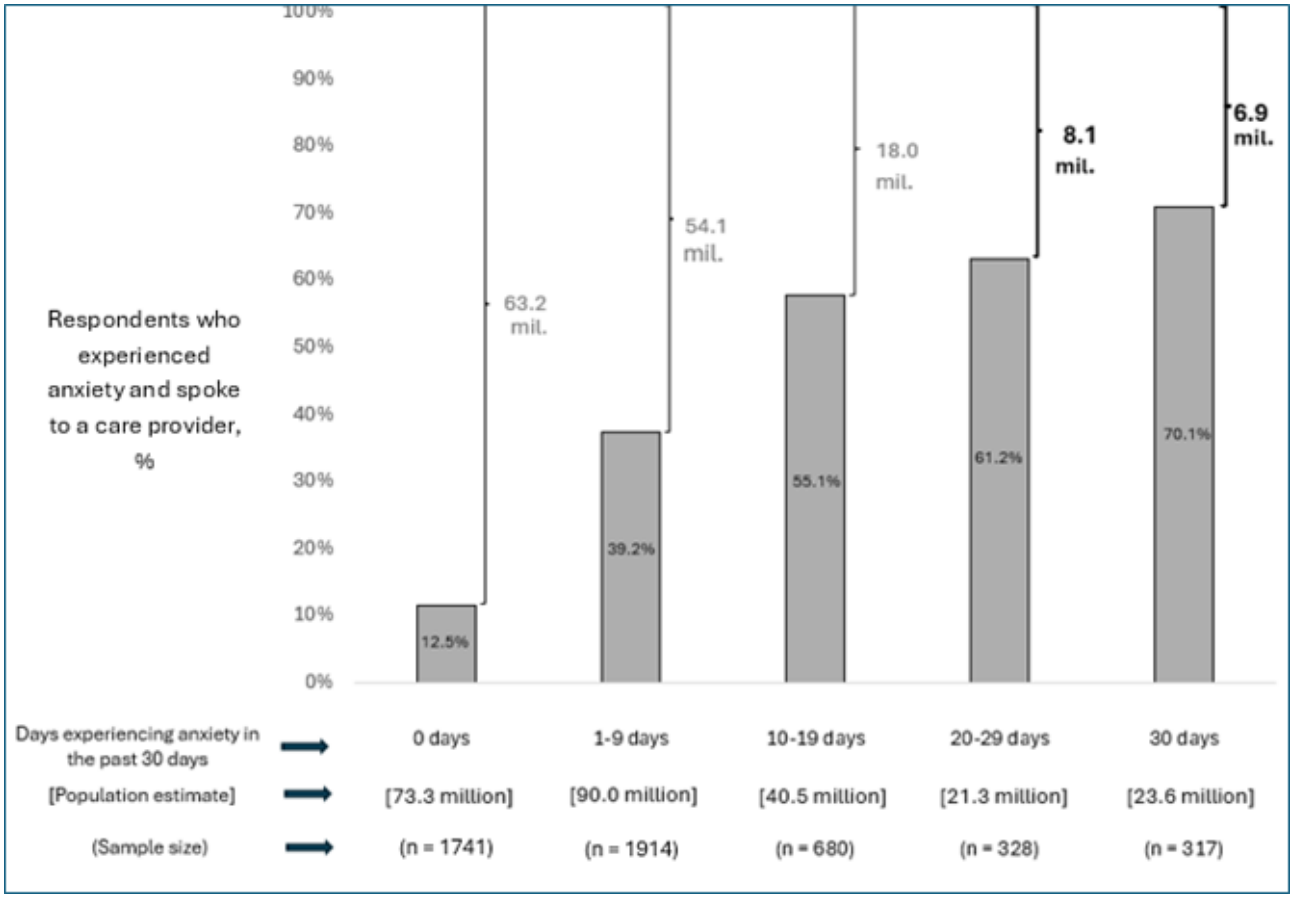
Figure 1. An estimated 10.2 million adults experienced feelings of depression almost every day but had never spoken to a health care provider about it. The figure plots the percentage of people who reported ever talking to a health care provider about depression on the y-axis against the number of days (category) they self-reported experiences of depression. The estimated population size and sample size, broken down across number of days reported (categorical variable), are displayed below the x-axis. The population estimates at the top of the figure, to the right of the curly brackets, represent the population estimates of the number of people who reported experiences of depression by number of days (categorical variable) but never talked to their health care provider about it. Percentages may not add up to 100 as some respondents may not have responded to one or both questions. Full details are available in Multimedia Appendix 1.



Additionally, we estimated that 44.9 million adults experienced anxiety on 20 or more days in the past month. Of them, an estimated 15 million had never spoken to an HCP about it (Figure 2).

XSL•FO
RenderX

Figure 2. An estimated 15.0 million adults experienced feelings of anxiety almost every day but had never spoken to a health care provider about it. The figure plots the percentage of people who reported ever talking to a health care provider about anxiety on the y-axis against the number of days (category) they self-reported experiences of anxiety. The estimated population size and sample size, broken down across number of days reported (categorical variable), are displayed below the x-axis. The population estimates at the top of the figure, to the right of the curly brackets, represent the population estimates of the number of people who reported experiences of anxiety by number of days (categorical variable) but never talked to their health care provider about it. Percentages may not add up to 100 as some respondents may not have responded to one or both questions. Full details are available in Multimedia Appendix 1.



Discussion

Principal Findings

We found that an estimated 53.8 million people have near-daily experiences of depression, anxiety, or both, and an estimated 18.6 million of them have never spoken to an HCP about these experiences.

Limitations

Our findings are based on self-reported data, and we cannot assess if people (1) have a mental health diagnosis, (2) need professional care, or (3) have sought or received care. However, the findings provide useful population-level information about people's self-reported experiences with depression or anxiety even though the results are not prevalence estimates.

Comparison to Prior Work

The Substance Abuse and Mental Health Services Administration reports that 58.7 million adults had any mental illness in 2023 [7]. Based on the National Health Interview Survey, the US Centers for Disease Control and Prevention reports that 18.2% and 21.4% of adults experienced symptoms of anxiety and depression, respectively, in 2022. In comparison, we find that 53.8 million adults (18.5% of the adult population in 2024, as per author calculations) self-reported near-daily experiences of depression, anxiety, or both.

Future Directions

This paper shows that millions of US adults experience depression, anxiety, or both on a near-daily basis but have never sought care from an HCP. The magnitude of these self-reported experiences, while alarming, does not correspond to a clinical diagnosis. Future research should aim to better understand the clinical needs of these people and build on the recent work showing a high degree of unmet need for mental health care [8].

XSL•FO
RenderX

Acknowledgments

EM-F had full access to all the data from the survey and takes responsibility for the integrity and accuracy of the data analysis. No artificial intelligence was used in the creation of this letter.

Data Availability

The datasets generated or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Both YP and EM-F contributed equally to this letter and had full access to the survey and results.

Conflicts of Interest

YP is an employee of Gary and Mary West Health Institute. EM-F is an employee of Gary and Mary West Health Policy Center. Both organizations are committed to addressing the need for mental health care.

Multimedia Appendix 1
Sampling strategy, list of questions, and detailed tabulations.
[DOCX File, 112 KB - i-jmr_v14i1e70626_app1.docx]

References

1. Goodwin RD, Dierker LC, Wu M, Galea S, Hoven CW, Weinberger AH. Trends in U.S. depression prevalence from 2015 to 2020: the widening treatment gap. Am J Prev Med 2022 Nov;63(5):726-733. [doi: 10.1016/j.amepre.2022.05.014] [Medline: 36272761]
2. Terlizzi EP, Zablotsky B. Symptoms of anxiety and depression among adults: United States, 2019 and 2022. Natl Health Stat Report 2024 Nov 7(213):CS353885. [doi: 10.15620/cdc/64018] [Medline: 39591466]
3. What is depression? American Psychiatric Association. 2024. URL: https://www.psychiatry.org/patients-families/depression/what-is-depression [accessed 2025-03-26]
4. Worry and anxiety. US Centers for Disease Control and Prevention. URL: https://www.cdc.gov/howrightnow/emotion/worry/index.html [accessed 2025-03-26]
5. Mongelli F, Georgakopoulos P, Pato MT. Challenges and opportunities to meet the mental health needs of underserved and disenfranchised populations in the United States. Focus (Am Psychiatr Publ) 2020 Jan;18(1):16-24. [doi: 10.1176/appi.focus.20190028] [Medline: 32047393]
6. Stigma, prejudice and discrimination against people with mental illness. American Psychiatric Association. 2024. URL: https://www.psychiatry.org/patients-families/stigma-and-discrimination [accessed 2025-03-26]
7. Substance abuse and mental health services administration. US Department of Health and Human Services. 2023. URL: https://www.samhsa.gov/data/sites/default/files/reports/rpt47095/National%20Report/National%20Report/2023-nsduh-annual-national.pdf [accessed 2025-05-16]
8. Meiselbach MK, Ettman CK, Shen K, Castrucci BC, Galea S. Unmet need for mental health care is common across insurance market segments in the United States. Health Aff Sch 2024 Mar;2(3):qxae032. [doi: 10.1093/haschl/qxae032] [Medline: 38756925]

Abbreviations

HCP: health care provider

Edited by TDA Cardoso; submitted 27.12.24; peer-reviewed by C Brantner, R Chen; revised version received 05.03.25; accepted 06.03.25; published 16.04.25.

Please cite as:
Pathak Y, Makk-Frid E
Population Estimates of Self-Reported Depression and Anxiety in the US From a National Survey: Cross-Sectional Survey Study
Interact J Med Res 2025;14:e70626
URL: https://www.i-jmr.org/2025/1/e70626
doi:10.2196/70626

XSL•FO
RenderX

© Yuvraj Pathak, Elvira Makk-Frid. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 16.4.2025. 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

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

XSL•FO
RenderX